



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: C12N 15/00	A2	(11) International Publication Number: WO 00/31247
		(43) International Publication Date: 02 June 2000 (02.06.2000)
(21) International Application Number: PCT/US99/27438		Published
(22) International Filing Date: 19 November 1999 (19.11.1999)		
(30) Priority Data: 60/109,401 20 November 1998 (20.11.1998) US 60/119,386 10 February 1999 (10.02.1999) US 60/122,620 03 March 1999 (03.03.1999) US 60/130,560 22 April 1999 (22.04.1999) US		
(60) Parent Application or Grant KOSAN BIOSCIENCES, INC. [/]; (). JULIEN, Bryan [/]; (). KATZ, Leonard [/]; (). KHOSLA, Chaitan [/]; (). TANG, Li [/]; (). ZIERMANN, Rainer [/]; (). MURASHIGE, Kate, H.; ().		
(54) Title: RECOMBINANT METHODS AND MATERIALS FOR PRODUCING EPOTHILONE AND EPOTHILONE DERIVATIVES		
(54) Titre: MATIERES ET PROCEDES RECOMBINANTS DESTINES A LA PRODUCTION D'EPOTHILONE ET DE DERIVES D'EPOTHILONE		
(57) Abstract		
<p>Recombinant nucleic acids that encode all or a portion of the epothilone polyketide synthase (PKS) are used to express recombinant PKS genes in host cells for the production of epothilones, epothilone derivatives, and polyketides that are useful as cancer chemotherapeutics, fungicides, and immunosuppressants.</p>		
(57) Abrégé		
<p>Selon cette invention, des acides nucléiques recombinants qui codent pour la synthèse de polycétides d'épothilones (PKS), intégralement ou en partie, sont utilisés pour exprimer les gènes de la PKS recombinante dans des cellules hôtes à des fins de fabrication d'épothilones, de dérivés d'épothilones et de polycétides utiles en tant qu'agents chimiothérapeutiques destinés au traitement du cancer ou utilisés en tant que fongicides ou immunosuppresseurs.</p>		

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/00		A2	(11) International Publication Number: WO 00/31247
			(43) International Publication Date: 2 June 2000 (02.06.00)
(21) International Application Number: PCT/US99/27438		(81) Designated States: AL, AM, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GE, HR, HU, IL, IS, JP, KG, KP, KR, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, UZ, VN, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 19 November 1999 (19.11.99)		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(30) Priority Data: 60/109,401 20 November 1998 (20.11.98) US 60/119,386 10 February 1999 (10.02.99) US 60/122,620 3 March 1999 (03.03.99) US 60/130,560 22 April 1999 (22.04.99) US			
(71) Applicant: KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Drive, Hayward, CA 94545 (US).			
(72) Inventors: JULIEN, Bryan; 4514 Fleming Avenue, Oakland, CA 94619 (US). KATZ, Leonard; 22294 City Center Drive, Apartment 5104, Hayward, CA 95051 (US). KHOSLA, Chaitan; 740 La Para Avenue, Palo Alto, CA 94306 (US). TANG, Li; 574 Cutwater Lane, Foster City, CA 94404 (US). ZIERMANN, Rainer; 3908 Leona Street, San Mateo, CA 94403 (US).			
(74) Agents: MURASHIGE, Kate, H. et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).			
(54) Title: RECOMBINANT METHODS AND MATERIALS FOR PRODUCING EPOTHILONE AND EPOTHILONE DERIVATIVES			
(57) Abstract Recombinant nucleic acids that encode all or a portion of the epothilone polyketide synthase (PKS) are used to express recombinant PKS genes in host cells for the production of epothilones, epothilone derivatives, and polyketides that are useful as cancer chemotherapeutics, fungicides, and immunosuppressants.			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

Description

5

10

15

20

25

30

35

40

45

50

55

RECOMBINANT METHODS AND MATERIALS FOR PRODUCING
EPOTHILONE AND EPOTHILONE DERIVATIVES

Reference to Government Funding

This invention was supported in part by SBIR grant 1R43-CA79228-01. The U.S. government has certain rights in this invention.

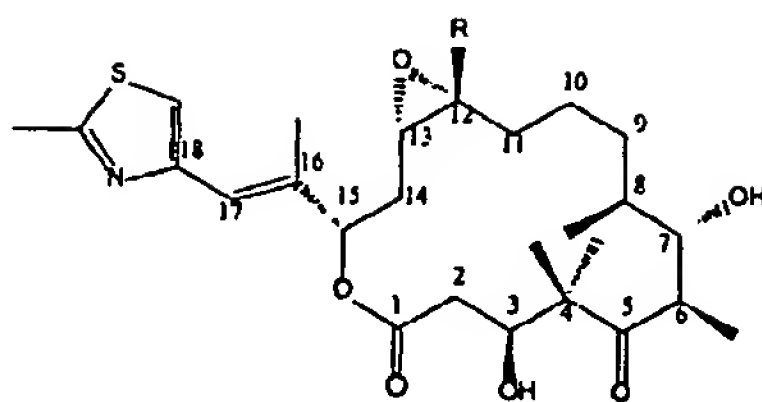
Field of the Invention

The present invention provides recombinant methods and materials for producing epothilone and epothilone derivatives. The invention relates to the fields of agriculture, chemistry, medicinal chemistry, medicine, molecular biology, and pharmacology.

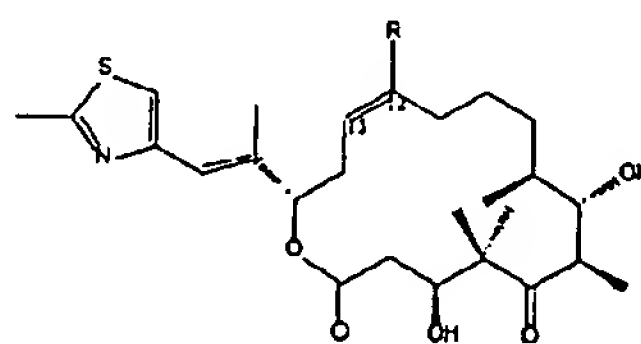
Background of the Invention

The epothilones were first identified by Gerhard Hofle and colleagues at the National Biotechnology Research Institute as an antifungal activity extracted from the myxobacterium *Sorangium cellulosum* (see K. Gerth *et al.*, 1996, J. Antibiotics 49: 560-563 and Germany Patent No. DE 41 38 042). The epothilones were later found to have activity in a tubulin polymerization assay (see D. Bollag *et al.*, 1995, Cancer Res. 55:2325-2333) to identify antitumor agents and have since been extensively studied as potential antitumor agents for the treatment of cancer.

The chemical structure of the epothilones produced by *Sorangium cellulosum* strain So ce 90 was described in Hofle *et al.*, 1996, Epothilone A and B - novel 16-membered macrolides with cytotoxic activity: isolation, crystal structure, and conformation in solution, Angew. Chem. Int. Ed. Engl. 35(13/14): 1567-1569, incorporated herein by reference. The strain was found to produce two epothilone compounds, designated A (R = H) and B (R = CH₃), as shown below, which showed broad cytotoxic activity against eukaryotic cells and noticeable activity and selectivity against breast and colon tumor cell lines.



The desoxy counterparts of epothilones A and B, also known as epothilones C (R = H) and D (R = CH₃), are known to be less cytotoxic, and the structures of these epothilones are shown below.



Two other naturally occurring epothilones have been described. These are epothilones E and F, in which the methyl side chain of the thiazole moiety of epothilones A and B has been hydroxylated to yield epothilones E and F, respectively.

Because of the potential for use of the epothilones as anticancer agents, and because of the low levels of epothilone produced by the native *So ce 90* strain, a number of research teams undertook the effort to synthesize the epothilones. This effort has been successful (see Balog *et al.*, 1996, Total synthesis of (-)-epothilone A, *Angew. Chem. Int. Ed. Engl.* 35(23/24): 2801-2803; Su *et al.*, 1997, Total synthesis of (-)-epothilone B: an extension of the Suzuki coupling method and insights into structure-activity relationships of the epothilones, *Angew. Chem. Int. Ed. Engl.* 36(7): 757-759; Meng *et al.*, 1997, Total syntheses of epothilones A and B, *JACS* 119(42): 10073-10092; and Balog *et al.*, 1998, A novel aldol condensation with 2-methyl-4-pentenal and its application to an improved total synthesis of epothilone B, *Angew. Chem. Int. Ed. Engl.* 37(19): 2675-2678, each of which is incorporated herein by reference). Despite the success of these efforts, the chemical synthesis of the epothilones is tedious, time-consuming, and expensive. Indeed, the methods have been characterized as impractical for the full-scale pharmaceutical development of an epothilone.

A number of epothilone derivatives, as well as epothilones A - D, have been studied *in vitro* and *in vivo* (see Su *et al.*, 1997, Structure-activity relationships of the epothilones and the first *in vivo* comparison with paclitaxel, *Angew. Chem. Int. Ed. Engl.*

5 36(19): 2093-2096; and Chou *et al.*, Aug. 1998, Desoxyepothilone B: an efficacious
microtubule-targeted antitumor agent with a promising *in vivo* profile relative to
epothilone B, Proc. Natl. Acad. Sci. USA 95: 9642-9647, each of which is incorporated
10 herein by reference). Additional epothilone derivatives and methods for synthesizing
5 epothilones and epothilone derivatives are described in PCT patent publication Nos.
99/54330, 99/54319, 99/54318, 99/43653, 99/43320, 99/42602, 99/40047, 99/27890,
99/07692, 99/02514, 99/01124, 98/25929, 98/22461, 98/08849, and 97/19086; U.S. Patent
15 No. 5,969,145; and Germany patent publication No. DE 41 38 042, each of which is
incorporated herein by reference.

10 There remains a need for economical means to produce not only the naturally
20 occurring epothilones but also the derivatives or precursors thereof, as well as new
epothilone derivatives with improved properties. There remains a need for a host cell that
produces epothilones or epothilone derivatives that is easier to manipulate and ferment
than the natural producer *Sorangium cellulosum*. The present invention meets these and
25 15 other needs.

Summary of the Invention

30 In one embodiment, the present invention provides recombinant DNA compounds
that encode the proteins required to produce epothilones A, B, C, and D. The present
20 invention also provides recombinant DNA compounds that encode portions of these
proteins. The present invention also provides recombinant DNA compounds that encode a
35 hybrid protein, which hybrid protein includes all or a portion of a protein involved in
epothilone biosynthesis and all or a portion of a protein involved in the biosynthesis of
another polyketide or non-ribosomal-derived peptide. In a preferred embodiment, the
25 recombinant DNA compounds of the invention are recombinant DNA cloning vectors that
40 facilitate manipulation of the coding sequences or recombinant DNA expression vectors
that code for the expression of one or more of the proteins of the invention in recombinant
host cells.

45 In another embodiment, the present invention provides recombinant host cells that
30 produce a desired epothilone or epothilone derivative. In one embodiment, the invention
provides host cells that produce one or more of the epothilones or epothilone derivatives at
higher levels than produced in the naturally occurring organisms that produce epothilones.
50 In another embodiment, the invention provides host cells that produce mixtures of

epothilones that are less complex than the mixtures produced by naturally occurring host cells. In another embodiment, the present invention provides non-*Sorangium* recombinant host cells that produce an epothilone or epothilone derivative.

In a preferred embodiment, the host cells of the invention produce less complex mixtures of epothilones than do naturally occurring cells that produce epothilones. Naturally occurring cells that produce epothilones typically produce a mixture of epothilones A, B, C, D, E, and F. The table below summarizes the epothilones produced in different illustrative host cells of the invention.

<u>Cell Type</u>	<u>Epothilones Produced</u>	<u>Epothilones Not Produced</u>
1	A, B, C, D, E, F	-----
2	A, C, E	B, D, F
3	B, D, F	A, C, E
4	A, B, C, D	E, F
5	A, C	B, D, E, F
6	C	A, B, D, E, F
7	B, D	A, C, E, F
8	D	A, B, C, E, F

In addition, cell types may be constructed which produce only the newly discovered epothilones G and H, further discussed below, and one or the other of G and H or both in combination with the downstream epothilones. Thus, it is understood, based on the present invention, that the biosynthetic pathway which relates the naturally occurring epothilones is, respectively, $G \rightarrow C \rightarrow A \rightarrow E$ and $H \rightarrow D \rightarrow B \rightarrow F$. Appropriate enzymes may also convert members of each pathway to the corresponding member of the other.

Thus, the recombinant host cells of the invention also include host cells that produce only one desired epothilone or epothilone derivative.

In another embodiment, the invention provides *Sorangium* host cells that have been modified genetically to produce epothilones either at levels greater than those observed in naturally occurring host cells or as less complex mixtures of epothilones than produced by naturally occurring host cells, or produce an epothilone derivative that is not produced in nature. In a preferred embodiment, the host cell produces the epothilones at equal to or greater than 20 mg/L.

5 In another embodiment, the recombinant host cells of the invention are host cells
other than *Sorangium cellulosum* that have been modified genetically to produce an
epothilone or an epothilone derivative. In a preferred embodiment, the host cell produces
10 the epothilones at equal to or greater than 20 mg/L. In a more preferred embodiment, the
5 recombinant host cells are *Myxococcus*, *Pseudomonas*, or *Streptomyces* host cells that
produce the epothilones or an epothilone derivative at equal to or greater than 20 mg/L.
15 In another embodiment, the present invention provides novel compounds useful in
agriculture, veterinary practice, and medicine. In one embodiment, the compounds are
useful as fungicides. In another embodiment, the compounds are useful in cancer
20 chemotherapy. In a preferred embodiment, the compound is an epothilone derivative that
is at least as potent against tumor cells as epothilone B or D. In another embodiment, the
compounds are useful as immunosuppressants. In another embodiment, the compounds are
useful in the manufacture of another compound. In a preferred embodiment, the
25 compounds are formulated in a mixture or solution for administration to a human or
15 animal.

These and other embodiments of the invention are described in more detail in the
following description, the examples, and claims set forth below.

Brief Description of the Figures

20 Figure 1 shows a restriction site map of the insert *Sorangium cellulosum* genomic
DNA in four overlapping cosmid clones (designated 8A3, 1A2, 4, and 85 and
35 corresponding to pKOS35-70.8A3, pKOS35-70.1A2, pKOS35-70.4, and pKOS35-79.85,
respectively) spanning the epothilone gene cluster. A functional map of the epothilone
gene cluster is also shown. The loading domain (Loading, *epoA*), the non-ribosomal
40 25 peptide synthase (NRPS, Module 1, *epoB*) module, and each module (Modules 2 through
9, *epoC*, *epoD*, *epoE*, and *epoF*) of the remaining eight modules of the epothilone synthase
gene are shown, as is the location of the *epoK* gene that encodes a cytochrome P450-like
epoxidation enzyme.

45 Figure 2 shows a number of precursor compounds to N-acylcysteamine thioester
30 derivatives that can be supplied to an epothilone PKS of the invention in which the NRPS-
like module 1 or module 2 KS domain has been inactivated to produce a novel epothilone
derivative. A general synthetic procedure for making such compounds is also shown.
50

5

Figure 3 shows restriction site and function maps of plasmids pKOS35-82.1 and pKOS35-82.2.

10

Figure 4 shows restriction site and function maps of plasmids pKOS35-154 and pKOS90-22.

5

Figure 5 shows a schematic of a protocol for introducing the epothilone PKS and modification enzyme genes into the chromosome of a *Myxococcus xanthus* host cell as described in Example 3.

15

Figure 6 shows restriction site and function maps of plasmids pKOS039-124 and pKOS039-124R.

10

Figure 7 shows a restriction site and function map of plasmid pKOS039-126R.

20

Figure 8 shows a restriction site and function map of plasmid pKOS039-141.

Figure 9 shows a restriction site and function map of plasmid pKOS045-12.

25

Detailed Description of the Invention

15

The present invention provides the genes and proteins that synthesize the epothilones in *Sorangium cellulosum* in recombinant and isolated form. As used herein, the term recombinant refers to a compound or composition produced by human intervention, typically by specific and directed manipulation of a gene or portion thereof. The term isolated refers to a compound or composition in a preparation that is substantially free of contaminating or undesired materials or, with respect to a compound or composition found in nature, substantially free of the materials with which that compound or composition is associated in its natural state. The epothilones (epothilone A, B, C, D, E, and F) and compounds structurally related thereto (epothilone derivatives) are potent cytotoxic agents specific for eukaryotic cells. These compounds have application as anti-fungals, cancer chemotherapeutics, and immunosuppressants. The epothilones are produced at very low levels in the naturally occurring *Sorangium cellulosum* cells in which they have been identified. Moreover, *S. cellulosum* is very slow growing, and fermentation of *S. cellulosum* strains is difficult and time-consuming. One important benefit conferred by the present invention is the ability simply to produce an epothilone or epothilone derivative in a non-*S. cellulosum* host cell. Another advantage of the present invention is the ability to produce the epothilones at higher levels and in greater amounts in the recombinant host cells provided by the invention than possible in the naturally

40

45

50

55

5 occurring epothilone producer cells. Yet another advantage is the ability to produce an epothilone derivative in a recombinant host cell.

10 The isolation of recombinant DNA encoding the epothilone biosynthetic genes resulted from the probing of a genomic library of *Sorangium cellulosum* SMP44 DNA. As
5 described more fully in Example 1 below, the library was prepared by partially digesting *S. cellulosum* genomic DNA with restriction enzyme SauIIA1 and inserting the DNA
15 fragments generated into BamHI-digested Supercos™ cosmid DNA (Stratagene). Cosmid clones containing epothilone gene sequences were identified by probing with DNA probes specific for sequences from PKS genes and reprobing with secondary probes comprising
10 nucleotide sequences identified with the primary probes.

20 Four overlapping cosmid clones were identified by this effort. These four cosmids were deposited with the American Type Culture Collection (ATCC), Manassas, VA, USA, under the terms of the Budapest Treaty, and assigned ATCC accession numbers. The
25 clones (and accession numbers) were designated as cosmids pKOS35-70.1A2 (ATCC 203782), pKOS35-70.4 (ATCC 203781), pKOS35-70.8A3 (ATCC 203783), and pKOS35-
15 79.85 (ATCC 203780). The cosmids contain insert DNA that completely spans the epothilone gene cluster. A restriction site map of these cosmids is shown in Figure 1.
30 Figure 1 also provides a function map of the epothilone gene cluster, showing the location of the six epothilone PKS genes and the *epoK* P450 epoxidase gene.

20 The epothilone PKS genes, like other PKS genes, are composed of coding sequences organized to encode a loading domain, a number of modules, and a thioesterase
35 domain. As described more fully below, each of these domains and modules corresponds to a polypeptide with one or more specific functions. Generally, the loading domain is responsible for binding the first building block used to synthesize the polyketide and
40 transferring it to the first module. The building blocks used to form complex polyketides are typically acylthioesters, most commonly acetyl, propionyl, malonyl, methylmalonyl, and ethylmalonyl CoA. Other building blocks include amino acid-like acylthioesters.
45 PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between the acylthioester building blocks. Each module is responsible for
30 binding a building block, performing one or more functions on that building block, and transferring the resulting compound to the next module. The next module, in turn, is
50 responsible for attaching the next building block and transferring the growing compound

5 to the next module until synthesis is complete. At that point, an enzymatic thioesterase (TE) activity cleaves the polyketide from the PKS.

10 Such modular organization is characteristic of the class of PKS enzymes that synthesize complex polyketides and is well known in the art. Recombinant methods for
5 manipulating modular PKS genes are described in U.S. Patent Nos. 5,672,491; 5,712,146; 5,830,750; and 5,843,718; and in PCT patent publication Nos. 98/49315 and 97/02358, each of which is incorporated herein by reference. The polyketide known as 6-
15 deoxyerythronolide B (6-dEB) is synthesized by a PKS that is a prototypical modular PKS enzyme. The genes, known as *eryAI*, *eryAII*, and *eryAIII*, that code for the multi-subunit
10 protein known as deoxyerythronolide B synthase or DEBS (each subunit is known as DEBS1, DEBS2, or DEBS3) that synthesizes 6-dEB are described in U.S. Patent Nos.
20 5,712,146 and 5,824,513, incorporated herein by reference.

The loading domain of the DEBS PKS consists of an acyltransferase (AT) and an acyl carrier protein (ACP). The AT of the DEBS loading domain recognizes propionyl
25 CoA (other loading domain ATs can recognize other acyl-CoAs, such as acetyl, malonyl, methylmalonyl, or butyryl CoA) and transfers it as a thioester to the ACP of the loading
15 domain. Concurrently, the AT on each of the six extender modules recognizes a methylmalonyl CoA (other extender module ATs can recognize other CoAs, such as
30 malonyl or alpha-substituted malonyl CoAs, i.e., malonyl, ethylmalonyl, and 2-
20 hydroxymalonyl CoA) and transfers it to the ACP of that module to form a thioester. Once DEBS is primed with acyl- and methylmalonyl-ACPs, the acyl group of the loading
35 domain migrates to form a thioester (trans-esterification) at the KS of the first module; at this stage, module one possesses an acyl-KS adjacent to a methylmalonyl ACP. The acyl
group derived from the DEBS loading domain is then covalently attached to the alpha-
25 carbon of the extender group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer
40 than the loading unit (elongation or extension). The growing polyketide chain is transferred from the ACP to the KS of the next module of DEBS, and the process
45 continues.

30 The polyketide chain, growing by two carbons for each module of DEBS, is sequentially passed as a covalently bound thioester from module to module, in an
assembly line-like process. The carbon chain produced by this process alone would
50 possess a ketone at every other carbon atom, producing a polyketone, from which the

5 name polyketide arises. Commonly, however, additional enzymatic activities modify the
beta keto group of each two carbon unit just after it has been added to the growing
polyketide chain but before it is transferred to the next module. Thus, in addition to the
10 minimal module containing KS, AT, and ACP necessary to form the carbon-carbon bond,
5 modules may contain a ketoreductase (KR) that reduces the keto group to an alcohol.
Modules may also contain a KR plus a dehydratase (DH) that dehydrates the alcohol to a
double bond. Modules may also contain a KR, a DH, and an enoylreductase (ER) that
15 converts the double bond to a saturated single bond using the beta carbon as a methylene
function. The DEBS modules include those with only a KR domain, only an inactive KR
10 domain, and with all three KR, DH, and ER domains.

20 Once a polyketide chain traverses the final module of a PKS, it encounters the
releasing domain or thioesterase found at the carboxyl end of most PKSs. Here, the
polyketide is cleaved from the enzyme and, for most but not all polyketides, cyclized. The
polyketide can be modified further by tailoring or modification enzymes; these enzymes
25 15 add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or
reduction, on the polyketide core molecule. For example, 6-dEB is hydroxylated,
methylated, and glycosylated (glycosidated) to yield the well known antibiotic
erythromycin A in the *Saccharopolyspora erythraea* cells in which it is produced
30 naturally.

20 While the above description applies generally to modular PKS enzymes and
specifically to DEBS, there are a number of variations that exist in nature. For example,
35 many PKS enzymes comprise loading domains that, unlike the loading domain of DEBS,
comprise an "inactive" KS domain that functions as a decarboxylase. This inactive KS is
in most instances called KS^Q, where the superscript is the single-letter abbreviation for the
25 amino acid (glutamine) that is present instead of the active site cysteine required for
ketosynthase activity. The epothilone PKS loading domain contains a KS^Y domain not
present in other PKS enzymes for which amino acid sequence is currently available in
which the amino acid tyrosine has replaced the cysteine. The present invention provides
45 recombinant DNA coding sequences for this novel KS domain.

30 Another important variation in PKS enzymes relates to the type of building block
incorporated. Some polyketides, including epothilone, incorporate an amino acid derived
building block. PKS enzymes that make such polyketides require specialized modules for
50 incorporation. Such modules are called non-ribosomal peptide synthetase (NRPS)

5 modules. The epothilone PKS, for example, contains an NRPS module. Another example of a variation relates to additional activities in a module. For example, one module of the epothilone PKS contains a methyltransferase (MT) domain, a heretofore unknown domain of PKS enzymes that make modular polyketides.

10 5 The complete nucleotide sequence of the coding sequence of the open reading frames (ORFs) of the epothilone PKS genes and epothilone tailoring (modification) enzyme genes is provided in Example 1, below. This sequence information together with the information provided below regarding the locations of the open reading frames of the genes within that sequence provides the amino acid sequence of the encoded proteins.

15 10 Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the epothilone PKS and epothilone modification enzymes of *Sorangium cellulosum* is shown herein merely to illustrate a preferred embodiment of the invention. The present

20 15 invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity and, in some instances, even an improvement of a desired activity. The present invention

25 20 includes such polypeptides with alternate amino acid sequences, and the amino acid sequences shown merely illustrate preferred embodiments of the invention.

30 35 The present invention provides recombinant genes for the production of epothilones. The invention is exemplified by the cloning, characterization, and manipulation of the epothilone PKS and modification enzymes of *Sorangium cellulosum* SMP44. The description of the invention and the recombinant vectors deposited in connection with that description enable the identification, cloning, and manipulation of epothilone PKS and modification enzymes from any naturally occurring host cell that produces an epothilone. Such host cells include other *S. cellulosum* strains, such as So ce 90, other *Sorangium* species, and non-*Sorangium* cells. Such identification, cloning, and

40 45 30 characterization can be conducted by those of ordinary skill in accordance with the present invention using standard methodology for identifying homologous DNA sequences and for identifying genes that encode a protein of function similar to a known protein. Moreover, the present invention provides recombinant epothilone PKS and modification

50 55

5 enzyme genes that are synthesized de novo or are assembled from non-epothilone PKS genes to provide an ordered array of domains and modules in one or more proteins that assemble to form a PKS that produces epothilone or an epothilone derivative.

10 The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and methods provided thereby, the following discussion describes various regions of the epothilone PKS and corresponding coding sequences. This discussion begins with a
15 general discussion of the genes that encode the PKS, the location of the various domains and modules in those genes, and the location of the various domains in those modules.
20 Then, a more detailed discussion follows, focusing first on the loading domain, followed by the NRPS module, and then the remaining eight modules of the epothilone PKS.

There are six epothilone PKS genes. The *epoA* gene encodes the 149 kDa loading domain (which can also be referred to as a loading module). The *epoB* gene encodes
25 module 1, the 158 kDa NRPS module. The *epoC* gene encodes the 193 kDa module 2. The *epoD* gene encodes a 765 kDa protein that comprises modules 3 through 6, inclusive. The *epoE* gene encodes a 405 kDa protein that comprises modules 7 and 8. The *epoF* gene encodes a 257 kDa protein that comprises module 9 and the thioesterase domain.
30 Immediately downstream of the *epoF* gene is *epoK*, the P450 epoxidase gene which encodes a 47 kDa protein, followed immediately by the *epoL* gene, which may encode a
20 24 kDa dehydratase. The *epoL* gene is followed by a number of ORFs that include genes believed to encode proteins involved in transport and regulation.

35 The sequences of these genes are shown in Example 1 in one contiguous sequence or contig of 71,989 nucleotides. This contig also contains two genes that appear to originate from a transposon and are identified below as ORF A and ORF B. These two
40 25 genes are believed not to be involved in epothilone biosynthesis but could possibly contain sequences that function as a promoter or enhancer. The contig also contains more than 12 additional ORFs, only 12 of which, designated ORF2 through ORF12 and ORF2 complement, are identified below. As noted, ORF2 actually is two ORFs, because the
45 complement of the strand shown also comprises an ORF. The function of the
30 corresponding gene product, if any, of these ORFs has not yet been established. The Table below provides the location of various open reading frames, module-coding sequences, and domain encoding sequences within the contig sequence shown in Example 1. Those of
50 skill in the art will recognize, upon consideration of the sequence shown in Example 1,

5

that the actual start locations of several of the genes could differ from the start locations shown in the table, because of the presence in frame codons for methionine or valine in close proximity to the codon indicated as the start codon. The actual start codon can be

10

confirmed by amino acid sequencing of the proteins expressed from the genes.

	<u>Start</u>	<u>Stop</u>	<u>Comment</u>
	3	992	transposase gene ORF A, not part of the PKS
	989	1501	transposase gene ORF B, not part of the PKS
15	1998	6263	<i>epoA</i> gene, encodes the loading domain
	2031	3548	KS ^Y of the loading domain
	3621	4661	AT of the loading domain
	4917	5810	ER of the loading domain, potentially involved in formation of the thiazole moiety
20	5856	6155	ACP of the loading domain
	6260	10493	<i>epoB</i> gene, encodes module 1, the NRPS module
	6620	6649	condensation domain C2 of the NRPS module
25	6861	6887	heterocyclization signature sequence
	6962	6982	condensation domain C4 of the NRPS module
	7358	7366	condensation domain C7 (partial) of the NRPS module
	7898	7921	adenylation domain A1 of the NRPS module
30	8261	8308	adenylation domain A3 of the NRPS module
	8411	8422	adenylation domain A4 of the NRPS module
	8861	8905	adenylation domain A6 of the NRPS module
	8966	8983	adenylation domain A7 of the NRPS module
35	9090	9179	adenylation domain A8 of the NRPS module
	9183	9992	oxidation region for forming thiazole
	10121	10138	Adenylation domain A10 of the NRPS module
	10261	10306	Thiolation domain (PCP) of the NRPS module
40	10639	16137	<i>epoC</i> gene, encodes module 2
	10654	12033	KS2, the KS domain of module 2
	12250	13287	AT2, the AT domain of module 2
	13327	13899	DH2, the DH domain of module 2
45	14962	15756	KR2, the KR domain of module 2
	15763	16008	ACP2, the ACP domain of module 2
	16134	37907	<i>epoD</i> gene, encodes modules 3-6
	16425	17606	KS3
50	17817	18857	AT3

55

5		<u>Start</u>	<u>Stop</u>	<u>Comment</u>
		19581	20396	KR3
		20424	20642	ACP3
		20706	22082	KS4
10		22296	23336	AT4
		24069	24647	KR4
		24867	25151	ACP4
		25203	26576	KS5
15		26793	27833	AT5
		27966	28574	DH5
		29433	30287	ER5
		30321	30869	KR5
20		31077	31373	ACP5
		31440	32807	KS6
		33018	34067	AT6
		34107	34676	DH6
25		35760	36641	ER6
		36705	37256	KR6
		37470	37769	ACP6
		37912	49308	<i>epoE</i> gene, encodes modules 7 and 8
30		38014	39375	KS7
		39589	40626	AT7
		41341	41922	KR7
		42181	42423	ACP7
35		42478	43851	KS8
		44065	45102	AT8
		45262	45810	DH (inactive)
		46072	47172	MT8, the methyltransferase domain of module 8
40		48103	48636	KR8, this domain is inactive
		48850	49149	ACP8
		49323	56642	<i>epoF</i> gene, encodes module 9 and the TE domain
		49416	50774	KS9
45		50985	52025	AT9
		52173	53414	DH (inactive)
		54747	55313	KR9
		55593	55805	ACP9
50		55878	56600	TE9, the thioesterase domain
		56757	58016	<i>epoK</i> gene, encodes the P450 epoxidase

5

10

15

20

<u>Start</u>	<u>Stop</u>	<u>Comment</u>
58194	58733	<i>epoL</i> gene (putative dehydratase)
59405	59974	ORF2 complement, complement of strand shown
59460	60249	ORF2
60271	60738	ORF3, complement of strand shown
61730	62647	ORF4 (putative transporter)
63725	64333	ORF5
64372	65643	ORF6
66237	67472	ORF7 (putative oxidoreductase)
67572	68837	ORF8 (putative oxidoreductase membrane subunit)
68837	69373	ORF9
69993	71174	ORF10 (putative transporter)
71171	71542	ORF11
71557	71989	ORF12

25

With this overview of the organization and sequence of the epothilone gene cluster, one can better appreciate the many different recombinant DNA compounds provided by the present invention.

30

35

40

45

50

55

5 The epothilone PKS is multiprotein complex composed of the gene products of the *epoA*, *epoB*, *epoC*, *epoD*, *epoE*, and *epoF* genes. To confer the ability to produce epothilones to a host cell, one provides the host cell with the recombinant *epoA*, *epoB*, *epoC*, *epoD*, *epoE*, and *epoF* genes of the present invention, and optionally other genes, capable of expression in that host cell. Those of skill in the art will appreciate that, while 10 the epothilone and other PKS enzymes may be referred to as a single entity herein, these enzymes are typically multisubunit proteins. Thus, one can make a derivative PKS (a PKS that differs from a naturally occurring PKS by deletion or mutation) or hybrid PKS (a PKS that is composed of portions of two different PKS enzymes) by altering one or more genes that encode one or more of the multiple proteins that constitute the PKS.

15 The post-PKS modification or tailoring of epothilone includes multiple steps mediated by multiple enzymes. These enzymes are referred to herein as tailoring or modification enzymes. Surprisingly, the products of the domains of the epothilone PKS predicted to be functional by analysis of the genes that encode them are compounds that have not been previously reported. These compounds are referred to herein as epothilones 20 G and H. Epothilones G and H lack the C-12-C-13 π -bond of epothilones C and D and the C-12-C-13 epoxide of epothilones A and B, having instead a hydrogen and hydroxyl

5 group at C-13, a single bond between C-12 and C-13, and a hydrogen and H or methyl group at C-12. These compounds are predicted to result from the epothilone PKS, because the DNA and corresponding amino acid sequence for module 4 of the epothilone PKS
10 does not appear to include a DH domain.

5 As described below, however, expression of the epothilone PKS genes *epoA*, *epoB*, *epoC*, *epoD*, *epoE*, and *epoF* in certain heterologous host cells that do not express *epoK* or *epoL* leads to the production of epothilones C and D, which lack the C-13 hydroxyl and
15 have a double bond between C-12 and C-13. The dehydration reaction that mediates the formation of this double bond may be due to the action of an as yet unrecognized domain of the epothilone PKS (for example, dehydration could occur in the next module, which
10 possesses an active DH domain and could generate a conjugated diene precursor prior to its dehydrogenation by an ER domain) or an endogenous enzyme in the heterologous host cells (*Streptomyces coelicolor*) in which it was observed. In the latter event, epothilones G and H may be produced in *Sorangium cellulosum* or other host cells and, to be converted
20 to epothilones C and D, by the action of a dehydratase, which may be encoded by the *epoL* gene. In any event, epothilones C and D are converted to epothilones A and B by an epoxidase encoded by the *epoK* gene. Epothilones A and B are converted to epothilones E and F by a hydroxylase gene, which may be encoded by one of the ORFs identified above or by another gene endogenous to *Sorangium cellulosum*. Thus, one can produce an
25 epothilone or epothilone derivative modified as desired in a host cell by providing that host cell with one or more of the recombinant modification enzyme genes provided by the invention or by utilizing a host cell that naturally expresses (or does not express) the modification enzyme. Thus, in general, by utilizing the appropriate host and by appropriate inactivation, if desired, of modification enzymes, one may interrupt the
30 progression of $G \rightarrow C \rightarrow A \rightarrow E$ or the corresponding downstream processing of epothilone H at any desired point; by controlling methylation, one or both of the pathways can be selected.

45 Thus, the present invention provides a wide variety of recombinant DNA compounds and host cells for expressing the naturally occurring epothilones A, B, C, and D and derivatives thereof. The invention also provides recombinant host cells, particularly
30 *Sorangium cellulosum* host cells that produce epothilone derivatives modified in a manner similar to epothilones E and F. Moreover, the invention provides host cells that can produce the heretofore unknown epothilones G and H, either by expression of the
50

5 epothilone PKS genes in host cells that do not express the dehydratase that converts
 epothilones G and H to C and D or by mutating or altering the PKS to abolish the
 dehydratase function, if it is present in the epothilone PKS.

10 The macrolide compounds that are products of the PKS cluster can thus be
5 modified in various ways. In addition to the modifications described above, the PKS
 products can be glycosylated, hydroxylated, dehydroxylated, oxidized, methylated and
15 demethylated using appropriate enzymes. Thus, in addition to modifying the product of
 the PKS cluster by altering the number, functionality, or specificity of the modules
 contained in the PKS, additional compounds within the scope of the invention can be
10 produced by additional enzyme-catalyzed activity either provided by a host cell in which
 the polyketide synthases are produced or by modifying these cells to contain additional
20 enzymes or by additional *in vitro* modification using purified enzymes or crude extracts
 or, indeed, by chemical modification.

25 The present invention also provides a wide variety of recombinant DNA
15 compounds and host cells that make epothilone derivatives. As used herein, the phrase
 "epothilone derivative" refers to a compound that is produced by a recombinant epothilone
 PKS in which at least one domain has been either rendered inactive, mutated to alter its
30 catalytic function, or replaced by a domain with a different function or in which a domain
 has been inserted. In any event, the "epothilone derivative PKS" functions to produce a
20 compound that differs in structure from a naturally occurring epothilone but retains its ring
 backbone structure and so is called an "epothilone derivative." To facilitate a better
35 understanding of the recombinant DNA compounds and host cells provided by the
 invention, a detailed discussion of the loading domain and each of the modules of the
 epothilone PKS, as well as novel recombinant derivatives thereof, is provided below.

40 The loading domain of the epothilone PKS includes an inactive KS domain, KS^Y,
 an AT domain specific for malonyl CoA (which is believed to be decarboxylated by the
 KS^Y domain to yield an acetyl group), and an ACP domain. The present invention
 provides recombinant DNA compounds that encode the epothilone loading domain. The
45 loading domain coding sequence is contained within an ~8.3 kb EcoRI restriction
30 fragment of cosmid pKOS35-70.8A3. The KS domain is referred to as inactive, because
 the active site region "TAYSSSL" of the KS domain of the loading domain has a Y
50 residue in place of the cysteine required for ketosynthase activity; this domain does have

5 decarboxylase activity. See Witkowski *et al.*, 7 Sep. 1999, Biochem. 38(36): 11643-11650, incorporated herein by reference.

10 The presence of the Y residue in place of a Q residue (which occurs typically in an inactive loading domain KS) may make the KS domain less efficient at decarboxylation.

5 The present invention provides a recombinant epothilone PKS loading domain and corresponding DNA sequences that encode an epothilone PKS loading domain in which the Y residue has been changed to a Q residue by changing the codon therefor in the
15 coding sequence of the loading domain. The present invention also provides recombinant PKS enzymes comprising such loading domains and host cells for producing such
10 enzymes and the polyketides produced thereby. These recombinant loading domains include those in which just the Y residue has been changed, those in which amino acids surrounding and including the Y domain have been changed, and those in which the
20 complete KS^Y domain has been replaced by a complete KS^Q domain. The latter embodiment includes but is not limited to a recombinant epothilone loading domain in
25 which the KS^Y domain has been replaced by the KS^Q domain of the oleandolide PKS or the narbonolide PKS (see the references cited below in connection with the oleandomycin, narbomycin, and picromycin PKS and modification enzymes).

30 The epothilone loading domain also contains an AT domain believed to bind malonyl CoA. The sequence "QTAFTQPALFTFEYALAALW...GHSIG" in the AT
20 domain is consistent with malonyl CoA specificity. As noted above, the malonyl CoA is believed to be decarboxylated by the KS^Y domain to yield acetyl CoA. The present
35 invention provides recombinant epothilone derivative loading domains or their encoding DNA sequences in which the malonyl specific AT domain or its encoding sequence has been changed to another specificity, such as methylmalonyl CoA, ethylmalonyl CoA, and
40 2-hydroxymalonyl CoA. When expressed with the other proteins of the epothilone PKS, such loading domains lead to the production of epothilones in which the methyl
substituent of the thiazole ring of epothilone is replaced with, respectively, ethyl, propyl, and hydroxymethyl. The present invention provides recombinant PKS enzymes
45 comprising such loading domains and host cells for producing such enzymes and the
30 polyketides produced thereby.

Those of skill in the art will recognize that an AT domain that is specific for 2-hydroxymalonyl CoA will result in a polyketide with a hydroxyl group at the
50 corresponding location in the polyketide produced, and that the hydroxyl group can be

5 methylated to yield a methoxy group by polyketide modification enzymes. See, e.g., the patent applications cited in connection with the FK-520 PKS in the table below. Consequently, reference to a PKS that has a 2-hydroxymalonyl specific AT domain herein
10 similarly refers to polyketides produced by that PKS that have either a hydroxyl or methoxyl group at the corresponding location in the polyketide.

The loading domain of the epothilone PKS also comprises an ER domain. While, this ER domain may be involved in forming one of the double bonds in the thiazole
15 moiety in epothilone (in the reverse of its normal reaction), or it may be non-functional. In either event, the invention provides recombinant DNA compounds that encode the epothilone PKS loading domain with and without the ER region, as well as hybrid loading
20 domains that contain an ER domain from another PKS (either active or inactive, with or without accompanying KR and DH domains) in place of the ER domain of the epothilone loading domain. The present invention also provides recombinant PKS enzymes comprising such loading domains and host cells for producing such enzymes and the
25 polyketides produced thereby.

The recombinant nucleic acid compounds of the invention that encode the loading domain of the epothilone PKS and the corresponding polypeptides encoded thereby are
30 useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the epothilone loading domain is coexpressed with the proteins of a heterologous PKS. As used herein, reference to a heterologous modular PKS (or to the coding sequence therefor) refers to all or part of a PKS, including each of the multiple
35 proteins constituting the PKS, that synthesizes a polyketide other than an epothilone or epothilone derivative (or to the coding sequences therefor). This coexpression can be in one of two forms. The epothilone loading domain can be coexpressed as a discrete protein
40 with the other proteins of the heterologous PKS or as a fusion protein in which the loading domain is fused to one or more modules of the heterologous PKS. In either event, the hybrid PKS formed, in which the loading domain of the heterologous PKS is replaced by the epothilone loading domain, provides a novel PKS. Examples of a heterologous PKS
45 that can be used to prepare such hybrid PKS enzymes of the invention include but are not limited to DEBS and the picromycin (narbonolide), oleandolide, rapamycin, FK-506, FK-
50 520, rifamycin, and avermectin PKS enzymes and their corresponding coding sequences.

In another embodiment, a nucleic acid compound comprising a sequence that encodes the epothilone loading domain is coexpressed with the proteins that constitute the

5 remainder of the epothilone PKS (i.e., the *epoB*, *epoC*, *epoD*, *epoE*, and *epoF* gene products) or a recombinant epothilone PKS that produces an epothilone derivative due to an alteration or mutation in one or more of the *epoB*, *epoC*, *epoD*, *epoE*, and *epoF* genes. As used herein, reference to an epothilone or a PKS that produces an epothilone derivative
10 (or to the coding sequence therefor) refers to all or any one of the proteins that comprise the PKS (or to the coding sequences therefor).

In another embodiment, the invention provides recombinant nucleic acid
15 compounds that encode a loading domain composed of part of the epothilone loading domain and part of a heterologous PKS. In this embodiment, the invention provides, for example, either replacing the malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT. This replacement, like the
20 others described herein, is typically mediated by replacing the coding sequences therefor to provide a recombinant DNA compound of the invention; the recombinant DNA is used to prepare the corresponding protein. Such changes (including not only replacements but
25 also deletions and insertions) may be referred to herein either at the DNA or protein level.

The compounds of the invention also include those in which both the KS^Y and AT domains of the epothilone loading domain have been replaced but the ACP and/or linker regions of the epothilone loading domain are left intact. Linker regions are those segments
30 of amino acids between domains in the loading domain and modules of a PKS that help form the tertiary structure of the protein and are involved in correct alignment and positioning of the domains of a PKS. These compounds include, for example, a recombinant loading domain coding sequence in which the KS^Y and AT domain coding
35 sequences of the epothilone PKS have been replaced by the coding sequences for the KS^Q and AT domains of, for example, the oleandolide PKS or the narbonolide PKS. There are also PKS enzymes that do not employ a KS^Q domain but instead merely utilize an AT
40 domain that binds acetyl CoA, propionyl CoA, or butyryl CoA (the DEBS loading domain) or isobutyryl CoA (the avermectin loading domain). Thus, the compounds of the invention also include, for example, a recombinant loading domain coding sequence in
45 which the KS^Y and AT domain coding sequences of the epothilone PKS have been replaced by an AT domain of the DEBS or avermectin PKS. The present invention also provides recombinant DNA compounds encoding loading domains in which the ACP
50 domain or any of the linker regions of the epothilone loading domain has been replaced by another ACP or linker region.

5

10

15

20

25

30

35

40

45

50

55

Any of the above loading domain coding sequences is coexpressed with the other proteins that constitute a PKS that synthesizes epothilone, an epothilone derivative, or another polyketide to provide a PKS of the invention. If the product desired is epothilone or an epothilone derivative, then the loading domain coding sequence is typically expressed as a discrete protein, as is the loading domain in the naturally occurring epothilone PKS. If the product desired is produced by the loading domain of the invention and proteins from one or more non-epothilone PKS enzymes, then the loading domain is expressed either as a discrete protein or as a fusion protein with one or more modules of the heterologous PKS.

The present invention also provides hybrid PKS enzymes in which the epothilone loading domain has been replaced in its entirety by a loading domain from a heterologous PKS with the remainder of the PKS proteins provided by modified or unmodified epothilone PKS proteins. The present invention also provides recombinant expression vectors and host cells for producing such enzymes and the polyketides produced thereby. In one embodiment, the heterologous loading domain is expressed as a discrete protein in a host cell that expresses the *epoB*, *epoC*, *epoD*, *epoE*, and *epoF* gene products. In another embodiment, the heterologous loading domain is expressed as a fusion protein with the *epoB* gene product in a host cell that expresses the *epoC*, *epoD*, *epoE*, and *epoF* gene products. In a related embodiment, the present invention provides recombinant epothilone PKS enzymes in which the loading domain has been deleted and replaced by an NRPS module and corresponding recombinant DNA compounds and expression vectors. In this embodiment, the recombinant PKS enzymes thus produce an epothilone derivative that comprises a dipeptide moiety, as in the compound leinamycin. The invention provides such enzymes in which the remainder of the epothilone PKS is identical in function to the native epothilone PKS as well as those in which the remainder is a recombinant PKS that produces an epothilone derivative of the invention.

The present invention also provides reagents and methods useful in deleting the loading domain coding sequence or any portion thereof from the chromosome of a host cell, such as *Sorangium cellulosum*, or replacing those sequences or any portion thereof with sequences encoding a recombinant loading domain. Using a recombinant vector that comprises DNA complementary to the DNA including and/or flanking the loading domain coding sequence in the *Sorangium* chromosome, one can employ the vector and

5 homologous recombination to replace the native loading domain coding sequence with a recombinant loading domain coding sequence or to delete the sequence altogether.

Moreover, while the above discussion focuses on deleting or replacing the
10 epothilone loading domain coding sequences, those of skill in the art will recognize that
5 the present invention provides recombinant DNA compounds, vectors, and methods useful
in deleting or replacing all or any portion of an epothilone PKS gene or an epothilone
modification enzyme gene. Such methods and materials are useful for a variety of
15 purposes. One purpose is to construct a host cell that does not make a naturally occurring
epothilone or epothilone derivative. For example, a host cell that has been modified to not
10 produce a naturally occurring epothilone may be particularly preferred for making
epothilone derivatives or other polyketides free of any naturally occurring epothilone.
20 Another purpose is to replace the deleted gene with a gene that has been altered so as to
provide a different product or to produce more of one product than another.

If the epothilone loading domain coding sequence has been deleted or otherwise
25 rendered non-functional in a *Sorangium cellulosum* host cell, then the resulting host cell
15 will produce a non-functional epothilone PKS. This PKS could still bind and process
extender units, but the thiazole moiety of epothilone would not form, leading to the
30 production of a novel epothilone derivative. Because this derivative would predictably
contain a free amino group, it would be produced at most in low quantities. As noted
20 above, however, provision of a heterologous or other recombinant loading domain to the
host cell would result in the production of an epothilone derivative with a structure
35 determined by the loading domain provided.

The loading domain of the epothilone PKS is followed by the first module of the
PKS, which is an NRPS module specific for cysteine. This NRPS module is naturally
25 expressed as a discrete protein, the product of the *epoB* gene. The present invention
40 provides the *epoB* gene in recombinant form. The recombinant nucleic acid compounds of
the invention that encode the NRPS module of the epothilone PKS and the corresponding
polypeptides encoded thereby are useful for a variety of applications. In one embodiment,
45 a nucleic acid compound comprising a sequence that encodes the epothilone NRPS
30 module is coexpressed with genes encoding one or more proteins of a heterologous PKS.
The NRPS module can be expressed as a discrete protein or as a fusion protein with one of
the proteins of the heterologous PKS. The resulting PKS, in which at least a module of the
50 heterologous PKS is replaced by the epothilone NRPS module or the NRPS module is in

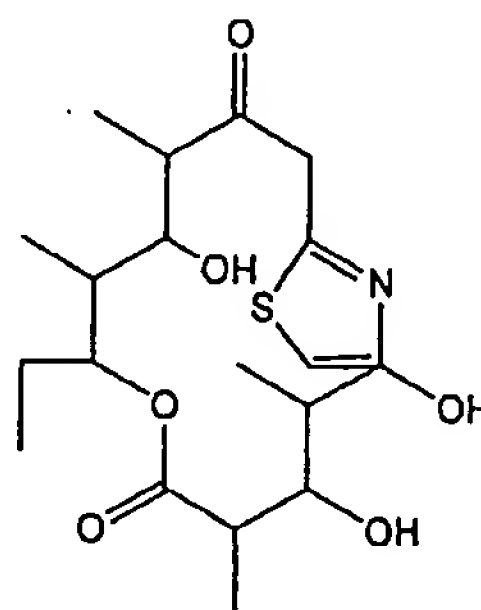
effect added as a module to the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the epothilone NRPS module is coexpressed with the other epothilone PKS proteins or modified versions thereof to provide a recombinant epothilone PKS that produces an epothilone or an epothilone derivative.

Two hybrid PKS enzymes provided by the invention illustrate this aspect. Both hybrid PKS enzymes are hybrids of DEBS and the epothilone NRPS module. The first hybrid PKS is composed of four proteins: (i) DEBS1; (ii) a fusion protein composed of the KS domain of module 3 of DEBS and all but the KS domain of the loading domain of the epothilone PKS; (iii) the epothilone NRPS module; and (iv) a fusion protein composed of the KS domain of module 2 of the epothilone PKS fused to the AT domain of module 5 of DEBS and the rest of DEBS3. This hybrid PKS produces a novel polyketide with a thiazole moiety incorporated into the macrolactone ring and a molecular weight of 413.53 when expressed in *Streptomyces coelicolor*. Glycosylated, hydroxylated, and methylated derivatives can be produced by expression of the hybrid PKS in *Saccharopolyspora erythraea*.

Diagrammatically, the construct is represented:



The structure of the product is:



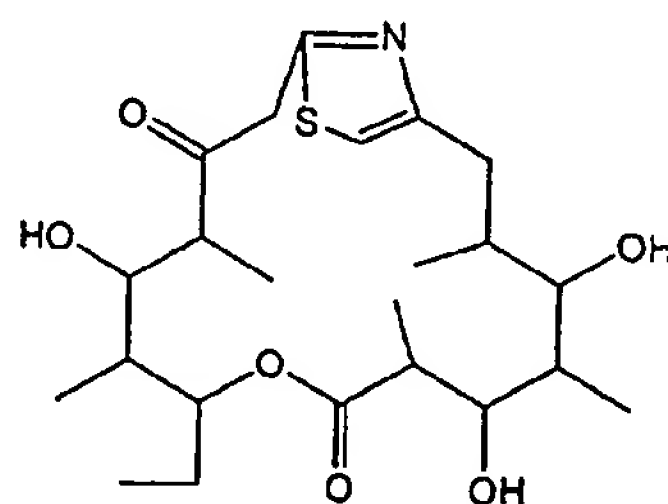
The second hybrid PKS illustrating this aspect of the invention is composed of five proteins: (i) DEBS1; (ii) a fusion protein composed of the KS domain of module 3 of DEBS and all but the KS domain of the loading domain of the epothilone PKS; (iii) the epothilone NRPS module; and (iv) a fusion protein composed of the KS domain of module

2 of the epothilone PKS fused to the AT domain of module 4 of DEBS and the rest of DEBS2; and (v) DEBS3. This hybrid PKS produces a novel polyketide with a thiazole moiety incorporated into the macrolactone ring and a molecular weight of 455.61 when expressed in *Streptomyces coelicolor*. Glycosylated, hydroxylated, and methylated derivatives can be produced by expression of the hybrid PKS in *Saccharopolyspora erythraea*.

Diagrammatically, the construct is represented:



The structure of the product is:



In another embodiment, a portion of the NRPS module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, changing the specificity of the NRPS module of the epothilone PKS from a cysteine to another amino acid. This change is accomplished by constructing a coding sequence in which all or a portion of the epothilone PKS NRPS module coding sequences have been replaced by those coding for an NRPS module of a different specificity. In one illustrative embodiment, the specificity of the epothilone NRPS module is changed from cysteine to serine or threonine. When the thus modified NRPS module is expressed with the other proteins of the epothilone PKS, the recombinant PKS produces an epothilone derivative in which the thiazole moiety of epothilone (or an epothilone derivative) is changed to an oxazole or 5-methyloxazole moiety, respectively. Alternatively, the present invention provides recombinant PKS enzymes composed of the products of the *epoA*, *epoC*, *epoD*, *epoE*, and *epoF* genes (or modified versions thereof) without an NRPS module or with an NRPS module from a heterologous PKS. The heterologous NRPS module can be expressed as a discrete protein or as a fusion protein with either the *epoA* or *epoC* genes.

5 The invention also provides methods and reagents useful in changing the
specificity of a heterologous NRPS module from another amino acid to cysteine. This
change is accomplished by constructing a coding sequence in which the sequences that
10 determine the specificity of the heterologous NRPS module have been replaced by those
5 that specify cysteine from the epothilone NRPS module coding sequence. The resulting
heterologous NRPS module is typically coexpressed in conjunction with the proteins
constituting a heterologous PKS that synthesizes a polyketide other than epothilone or an
15 epothilone derivative, although the heterologous NRPS module can also be used to
produce epothilone or an epothilone derivative.

10 In another embodiment, the invention provides recombinant epothilone PKS
enzymes and corresponding recombinant nucleic acid compounds and vectors in which the
20 NRPS module has been inactivated or deleted. Such enzymes, compounds, and vectors are
constructed generally in accordance with the teaching for deleting or inactivating the
epothilone PKS or modification enzyme genes above. Inactive NRPS module proteins and
25 the coding sequences therefore provided by the invention include those in which the
peptidyl carrier protein (PCP) domain has been wholly or partially deleted or otherwise
rendered inactive by changing the active site serine (the site for phosphopantetheinylation)
30 to another amino acid, such as alanine, or the adenylation domains have been deleted or
otherwise rendered inactive. In one embodiment, both the loading domain and the NRPS
20 have been deleted or rendered inactive. In any event, the resulting epothilone PKS can
then function only if provided a substrate that binds to the KS domain of module 2 (or a
35 subsequent module) of the epothilone PKS or a PKS for an epothilone derivative. In a
method provided by the invention, the thus modified cells are then fed activated
acylthioesters that are bound by preferably the second, but potentially any subsequent,
25 module and processed into novel epothilone derivatives.

40 Thus, in one embodiment, the invention provides *Sorangium* and non-*Sorangium*
host cells that express an epothilone PKS (or a PKS that produces an epothilone
derivative) with an inactive NRPS. The host cell is fed activated acylthioesters to produce
45 novel epothilone derivatives of the invention. The host cells expressing, or cell free
30 extracts containing, the PKS can be fed or supplied with N-acylcysteamine thioesters
(NACS) of novel precursor molecules to prepare epothilone derivatives. See U.S. patent
application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT patent publication No.
50 US99/03986, both of which are incorporated herein by reference, and Example 6, below.

5 The second (first non-NRPS) module of the epothilone PKS includes a KS, an AT specific for methylmalonyl CoA, a DH, a KR, and an ACP. This module is encoded by a sequence within an ~13.1 kb EcoRI-NsiI restriction fragment of cosmid pKOS35-70.8A3.

10 The recombinant nucleic acid compounds of the invention that encode the second
5 module of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. The second module of the epothilone PKS is produced as a discrete protein by the *epoC* gene. The present invention provides the *epoC* gene in
15 recombinant form. In one embodiment, a DNA compound comprising a sequence that encodes the epothilone second module is coexpressed with the proteins constituting a
10 heterologous PKS either as a discrete protein or as a fusion protein with one or more modules of the heterologous PKS. The resulting PKS, in which a module of the
20 heterologous PKS is either replaced by the second module of the epothilone PKS or the latter is merely added to the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the second
25 module of the epothilone PKS is coexpressed with the other proteins constituting the epothilone PKS or a recombinant epothilone PKS that produces an epothilone derivative.

In another embodiment, all or only a portion of the second module coding
30 sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the
20 methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting either the DH or KR or both; replacing the DH
35 or KR or both with a DH or KR or both that specify a different stereochemistry; and/or inserting an ER. Generally, any reference herein to inserting or replacing a PKS KR, DH, and/or ER domain includes the replacement of the associated KR, DH, or ER domains in
25 that module, typically with corresponding domains from the module from which the inserted or replacing domain is obtained. In addition, the KS and/or ACP can be replaced
40 with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the epothilone PKS, from a gene for a PKS that produces a polyketide
45 other than epothilone, or from chemical synthesis. The resulting heterologous second module coding sequence can be coexpressed with the other proteins that constitute a PKS that synthesizes epothilone, an epothilone derivative, or another polyketide. Alternatively,
50 one can delete or replace the second module of the epothilone PKS with a module from a

5 heterologous PKS, which can be expressed as a discrete protein or as a fusion protein fused to either the *epoB* or *epoD* gene product.

10 Illustrative recombinant PKS genes of the invention include those in which the AT domain encoding sequences for the second module of the epothilone PKS have been
5 altered or replaced to change the AT domain encoded thereby from a methylmalonyl specific AT to a malonyl specific AT. Such malonyl specific AT domain encoding nucleic acids can be isolated, for example and without limitation, from the PKS genes encoding the narbonolide PKS, the rapamycin PKS (i.e., modules 2 and 12), and the FK-520 PKS (i.e., modules 3, 7, and 8). When such a hybrid second module is coexpressed with the
15 other proteins constituting the epothilone PKS, the resulting epothilone derivative produced is a 16-desmethyl epothilone derivative.

20 In addition, the invention provides DNA compounds and vectors encoding recombinant epothilone PKS enzymes and the corresponding recombinant proteins in which the KS domain of the second (or subsequent) module has been inactivated or
25 deleted. In a preferred embodiment, this inactivation is accomplished by changing the codon for the active site cysteine to an alanine codon. As with the corresponding variants described above for the NRPS module, the resulting recombinant epothilone PKS enzymes are unable to produce an epothilone or epothilone derivative unless supplied a precursor
30 that can be bound and extended by the remaining domains and modules of the recombinant PKS enzyme. Illustrative diketides are described in Example 6, below.

35 The third module of the epothilone PKS includes a KS, an AT specific for malonyl CoA, a KR, and an ACP. This module is encoded by a sequence within an ~8 kb BglII-NsiI restriction fragment of cosmid pKOS35-70.8A3.

40 The recombinant DNA compounds of the invention that encode the third module of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. The third module of the epothilone PKS is expressed in a protein, the product of the *epoD* gene, which also contains modules 4, 5, and 6. The present invention provides the *epoD* gene in recombinant form. The present invention also
45 provides recombinant DNA compounds that encode each of the epothilone PKS modules 3, 4, 5, and 6, as discrete coding sequences without coding sequences for the other
50 epothilone modules. In one embodiment, a DNA compound comprising a sequence that encodes the epothilone third module is coexpressed with proteins constituting a heterologous PKS. The third module of the epothilone PKS can be expressed either as a

5 discrete protein or as a fusion protein fused to one or more modules of the heterologous
PKS. The resulting PKS, in which a module of the heterologous PKS is either replaced by
that for the third module of the epothilone PKS or the latter is merely added to the
10 modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA
5 compound comprising a sequence that encodes the third module of the epothilone PKS is
coexpressed with proteins comprising the remainder of the epothilone PKS or a
recombinant epothilone PKS that produces an epothilone derivative, typically as a protein
15 comprising not only the third but also the fourth, fifth, and sixth modules.

In another embodiment, all or a portion of the third module coding sequence is
10 utilized in conjunction with other PKS coding sequences to create a hybrid module. In this
embodiment, the invention provides, for example, either replacing the malonyl CoA
specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA
specific AT; deleting the KR; replacing the KR with a KR that specifies a different
20 stereochemistry; and/or inserting a DH or a DH and an ER. As above, the reference to
15 inserting a DH or a DH and an ER includes the replacement of the KR with a DH and KR
or an ER, DH, and KR. In addition, the KS and/or ACP can be replaced with another KS
and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH,
KR, ER, or ACP coding sequence can originate from a coding sequence for another
30 module of the epothilone PKS, from a coding sequence for a PKS that produces a
polyketide other than epothilone, or from chemical synthesis. The resulting heterologous
20 third module coding sequence can be utilized in conjunction with a coding sequence for a
PKS that synthesizes epothilone, an epothilone derivative, or another polyketide.
35

Illustrative recombinant PKS genes of the invention include those in which the AT
domain encoding sequences for the third module of the epothilone PKS have been altered
25 or replaced to change the AT domain encoded thereby from a malonyl specific AT to a
methylmalonyl specific AT. Such methylmalonyl specific AT domain encoding nucleic
40 acids can be isolated, for example and without limitation, from the PKS genes encoding
DEBS, the narbonolide PKS, the rapamycin PKS, and the FK-520 PKS. When
45 coexpressed with the remaining modules and proteins of the epothilone PKS or an
30 epothilone PKS derivative, the recombinant PKS produces the 14-methyl epothilone
derivatives of the invention.

Those of skill in the art will recognize that the KR domain of the third module of
50 the PKS is responsible for forming the hydroxyl group involved in cyclization of

5 epothilone. Consequently, abolishing the KR domain of the third module or adding a DH or DH and ER domains will interfere with the cyclization, leading either to a linear molecule or to a molecule cyclized at a different location than is epothilone.

10 The fourth module of the epothilone PKS includes a KS, an AT that can bind either
5 malonyl CoA or methylmalonyl CoA, a KR, and an ACP. This module is encoded by a sequence within an ~10 kb NsiI-HindIII restriction fragment of cosmid pKOS35-70.1A2.

15 The recombinant DNA compounds of the invention that encode the fourth module of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence
10 that encodes the epothilone fourth module is inserted into a DNA compound that comprises the coding sequence for one or more modules of a heterologous PKS. The resulting construct encodes a protein in which a module of the heterologous PKS is either
20 replaced by that for the fourth module of the epothilone PKS or the latter is merely added to the modules of the heterologous PKS. Together with other proteins that constitute the
25 heterologous PKS, this protein provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fourth module of the epothilone PKS is expressed in a host cell that also expresses the remaining modules and proteins of the
30 epothilone PKS or a recombinant epothilone PKS that produces an epothilone derivative. For making epothilone or epothilone derivatives, the recombinant fourth module is usually
20 expressed in a protein that also contains the epothilone third, fifth, and sixth modules or modified versions thereof.

35 In another embodiment, all or a portion of the fourth module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the malonyl CoA and
25 methylmalonyl specific AT with a malonyl CoA, methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR; and/or replacing the KR, including, optionally, to specify a different stereochemistry; and/or inserting a DH or a DH
40 and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP
45 coding sequence can originate from a coding sequence for another module of the epothilone PKS, from a gene for a PKS that produces a polyketide other than epothilone, or from chemical synthesis. The resulting heterologous fourth module coding sequence is
30 incorporated into a protein subunit of a recombinant PKS that synthesizes epothilone, an
50

5 epothilone derivative, or another polyketide. If the desired polyketide is an epothilone or
epothilone derivative, the recombinant fourth module is typically expressed as a protein
that also contains the third, fifth, and sixth modules of the epothilone PKS or modified
10 versions thereof. Alternatively, the invention provides recombinant PKS enzymes for
5 epothilones and epothilone derivatives in which the entire fourth module has been deleted
or replaced by a module from a heterologous PKS.

15 In a preferred embodiment, the invention provides recombinant DNA compounds
comprising the coding sequence for the fourth module of the epothilone PKS modified to
encode an AT that binds methylmalonyl CoA and not malonyl CoA. These recombinant
10 molecules are used to express a protein that is a recombinant derivative of the *epoD*
protein that comprises the modified fourth module as well as modules 3, 5, and 6, any one
20 or more of which can optionally be in derivative form, of the epothilone PKS. In another
preferred embodiment, the invention provides recombinant DNA compounds comprising
the coding sequence for the fourth module of the epothilone PKS modified to encode an
25 15 AT that binds malonyl CoA and not methylmalonyl CoA. These recombinant molecules
are used to express a protein that is a recombinant derivative of the *epoD* protein that
comprises the modified fourth module as well as modules 3, 5, and 6, any one or more of
which can optionally be in derivative form, of the epothilone PKS.

30 Prior to the present invention, it was known that *Sorangium cellulosum* produced
20 epothilones A, B, C, D, E, and F and that epothilones A, C, and E had a hydrogen at C-12,
while epothilones B, D, and F had a methyl group at this position. Unappreciated prior to
35 the present invention was the order in which these compounds were synthesized in
S. cellulosum, and the mechanism by which some of the compounds had a hydrogen at C-
12 where others had a methyl group at this position. The present disclosure reveals that
25 epothilones A and B are derived from epothilones C and D by action of the *epoK* gene
40 product and that the presence of a hydrogen or methyl moiety at C-12 is due to the AT
domain of module 4 of the epothilone PKS. This domain can bind either malonyl or
methylmalonyl CoA and, consistent with its having greater similarity to malonyl specific
45 AT domains than to methylmalonyl specific AT domains, binds malonyl CoA more often
30 than methylmalonyl CoA.

Thus, the invention provides recombinant DNA compounds and expression vectors
50 and the corresponding recombinant PKS in which the hybrid fourth module with a
methylmalonyl specific AT has been incorporated. The methylmalonyl specific AT coding

5 sequence can originate, for example and without limitation, from coding sequences for the
oleandolide PKS, DEBS, the narbonolide PKS, the rapamycin PKS, or any other PKS that
comprises a methylmalonyl specific AT domain. In accordance with the invention, the
10 hybrid fourth module expressed from this coding sequence is incorporated into the
5 epothilone PKS (or the PKS for an epothilone derivative), typically as a derivative *epoD*
gene product. The resulting recombinant epothilone PKS produces epothilones with a
methyl moiety at C-12, i.e., epothilone H (or an epothilone H derivative) if there is no
15 dehydratase activity to form the C-12-C-13 alkene; epothilone D (or an epothilone D
derivative), if the dehydratase activity but not the epoxidase activity is present; epothilone
10 B (or an epothilone B derivative), if both the dehydratase and epoxidase activity but not
the hydroxylase activity are present; and epothilone F (or an epothilone F derivative), if all
20 three dehydratase, epoxidase, and hydroxylase activities are present. As indicated
parenthetically above, the cell will produce the corresponding epothilone derivative if
there have been other changes to the epothilone PKS.

25 15 If the recombinant PKS comprising the hybrid methylmalonyl specific fourth
module is expressed in, for example, *Sorangium cellulosum*, the appropriate modifying
enzymes are present (unless they have been rendered inactive in accordance with the
30 methods herein), and epothilones D, B, and/or F are produced. Such production is
typically carried out in a recombinant *S. cellulosum* provided by the present invention in
20 which the native epothilone PKS is unable to function at all or unable to function except in
conjunction with the recombinant fourth module provided. In an illustrative example, one
35 can use the methods and reagents of the invention to render inactive the *epoD* gene in the
native host. Then, one can transform that host with a vector comprising the recombinant
epoD gene containing the hybrid fourth module coding sequence. The recombinant vector
25 can exist as an extrachromosomal element or as a segment of DNA integrated into the host
40 cell chromosome. In the latter embodiment, the invention provides that one can simply
integrate the recombinant methylmalonyl specific module 4 coding sequence into wild-
type *S. cellulosum* by homologous recombination with the native *epoD* gene to ensure that
45 only the desired epothilone is produced. The invention provides that the *S. cellulosum* host
30 can either express or not express (by mutation or homologous recombination of the native
genes therefor) the dehydratase, epoxidase, and/or oxidase gene products and thus form or
not form the corresponding epothilone D, B, and F compounds, as the practitioner elects.
50

5 *Sorangium cellulosum* modified as described above is only one of the recombinant
host cells provided by the invention. In a preferred embodiment, the recombinant
methylmalonyl specific epothilone fourth module coding sequences are used in
10 accordance with the methods of invention to produce epothilone D, B, and F (or their
5 corresponding derivatives) in heterologous host cells. Thus, the invention provides
reagents and methods for introducing the epothilone or epothilone derivative PKS and
epothilone dehydratase, epoxidase, and hydroxylase genes and combinations thereof into
15 heterologous host cells.

 The recombinant methylmalonyl specific epothilone fourth module coding
10 sequences provided by the invention afford important alternative methods for producing
desired epothilone compounds in host cells. Thus, the invention provides a hybrid fourth
20 module coding sequence in which, in addition to the replacement of the endogenous AT
coding sequence with a coding sequence for an AT specific for methylmalonyl Co A,
coding sequences for a DH and KR for, for example and without limitation, module 10 of
25 15 the rapamycin PKS or modules 1 or 5 of the FK-520 PKS have replaced the endogenous
KR coding sequences. When the gene product comprising the hybrid fourth module and
epothilone PKS modules 3, 5, and 6 (or derivatives thereof) encoded by this coding
30 sequence is incorporated into a PKS comprising the other epothilone PKS proteins (or
derivatives thereof) produced in a host cell, the cell makes either epothilone D or its trans
20 stereoisomer (or derivatives thereof), depending on the stereochemical specificity of the
inserted DH and KR domains.

35 Similarly, and as noted above, the invention provides recombinant DNA
compounds comprising the coding sequence for the fourth module of the epothilone PKS
modified to encode an AT that binds malonyl CoA and not methylmalonyl CoA. The
25 invention provides recombinant DNA compounds and vectors and the corresponding
40 recombinant PKS in which this hybrid fourth module has been incorporated into a
derivative *epoD* gene product. When incorporated into the epothilone PKS (or the PKS for
an epothilone derivative), the resulting recombinant epothilone PKS produces epothilones
45 C, A, and E, depending, again, on whether epothilone modification enzymes are present.
30 As noted above, depending on the host, whether the fourth module includes a KR and DH
domain, and on whether and which of the dehydratase, epoxidase, and oxidase activities
are present, the practitioner of the invention can produce one or more of the epothilone G,
50

5

C, A, and E compounds and derivatives thereof using the compounds, host cells, and methods of the invention.

10

The fifth module of the epothilone PKS includes a KS, an AT that binds malonyl CoA, a DH, an ER, a KR, and an ACP. This module is encoded by a sequence within an
5 ~12.4 kb NsiI-NotI restriction fragment of cosmid pKOS35-70.1A2.

15

The recombinant DNA compounds of the invention that encode the fifth module of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the epothilone fifth module is inserted into a DNA compound that comprises the
10 coding sequence for one or more modules of a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fifth module of the epothilone PKS or the latter is merely added to coding
20 sequences for the modules of the heterologous PKS, can be incorporated into an expression vector and used to produce the recombinant protein encoded thereby. When the
25 recombinant protein is combined with the other proteins of the heterologous PKS, a novel PKS is produced. In another embodiment, a DNA compound comprising a sequence that encodes the fifth module of the epothilone PKS is inserted into a DNA compound that
30 comprises coding sequences for the epothilone PKS or a recombinant epothilone PKS that produces an epothilone derivative. In the latter constructs, the epothilone fifth module is
20 typically expressed as a protein comprising the third, fourth, and sixth modules of the epothilone PKS or derivatives thereof.

35

In another embodiment, a portion of the fifth module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module coding sequence and the hybrid module encoded thereby. In this embodiment, the invention provides, for
25 example, either replacing the malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one, two, or all three of the ER, DH, and KR; and/or replacing any one, two, or all three of the ER, DH, and KR with either a KR, a DH and KR, or a KR, DH, and ER, including, optionally, to
40 specify a different stereochemistry. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for
45 another module of the epothilone PKS, from a coding sequence for a PKS that produces a polyketide other than epothilone, or from chemical synthesis. The resulting hybrid fifth
30

50

55

5 module coding sequence can be utilized in conjunction with a coding sequence for a PKS
that synthesizes epothilone, an epothilone derivative, or another polyketide. Alternatively,
the fifth module of the epothilone PKS can be deleted or replaced in its entirety by a
10 module of a heterologous PKS to produce a protein that in combination with the other
5 proteins of the epothilone PKS or derivatives thereof constitutes a PKS that produces an
epothilone derivative.

15 Illustrative recombinant PKS genes of the invention include recombinant *epoD*
gene derivatives in which the AT domain encoding sequences for the fifth module of the
epothilone PKS have been altered or replaced to change the AT domain encoded thereby
10 from a malonyl specific AT to a methylmalonyl specific AT. Such methylmalonyl specific
AT domain encoding nucleic acids can be isolated, for example and without limitation,
20 from the PKS genes encoding DEBS, the narbonolide PKS, the rapamycin PKS, and the
FK-520 PKS. When such recombinant *epoD* gene derivatives are coexpressed with the
epoA, *epoB*, *epoC*, *epoE*, and *epoF* genes (or derivatives thereof), the PKS composed
25 thereof produces the 10-methyl epothilones or derivatives thereof. Another recombinant
epoD gene derivative provided by the invention includes not only this altered module 5
coding sequence but also module 4 coding sequences that encode an AT domain that binds
only methylmalonyl CoA. When incorporated into a PKS with the *epoA*, *epoB*, *epoC*,
30 *epoE*, and *epoF* genes, the recombinant *epoD* gene derivative product leads to the
20 production of 10-methyl epothilone B and/or D derivatives.

Other illustrative recombinant *epoD* gene derivatives of the invention include those
35 in which the ER, DH, and KR domain encoding sequences for the fifth module of the
epothilone PKS have been replaced with those encoding (i) a KR and DH domain; (ii) a
KR domain; and (iii) an inactive KR domain. These recombinant *epoD* gene derivatives of
25 the invention are coexpressed with the *epoA*, *epoB*, *epoC*, *epoE*, and *epoF* genes to
40 produce a recombinant PKS that makes the corresponding (i) C-11 alkene, (ii) C-11
hydroxy, and (iii) C-11 keto epothilone derivatives. These recombinant *epoD* gene
derivatives can also be coexpressed with recombinant *epo* genes containing other
45 alterations or can themselves be further altered to produce a PKS that makes the
30 corresponding C-11 epothilone derivatives. For example, one recombinant *epoD* gene
derivative provided by the invention also includes module 4 coding sequences that encode
an AT domain that binds only methylmalonyl CoA. When incorporated into a PKS with
50 the *epoA*, *epoB*, *epoC*, *epoE*, and *epoF* genes, the recombinant *epoD* gene derivative

5 product leads to the production of the corresponding C-11 epothilone B and/or D derivatives.

Functionally similar *epoD* genes for producing the epothilone C-11 derivatives can also be made by inactivation of one, two, or all three of the ER, DH, and KR domains of the epothilone fifth module. However, the preferred mode for altering such domains in any module is by replacement with the complete set of desired domains taken from another module of the same or a heterologous PKS coding sequence. In this manner, the natural architecture of the PKS is conserved. Also, when present, KR and DH or KR, DH, and ER domains that function together in a native PKS are preferably used in the recombinant PKS. Illustrative replacement domains for the substitutions described above include, for example and without limitation, the inactive KR domain from the rapamycin PKS module 3 to form the ketone, the KR domain from the rapamycin PKS module 5 to form the alcohol, and the KR and DH domains from the rapamycin PKS module 4 to form the alkene. Other such inactive KR, active KR, and active KR and DH domain encoding nucleic acids can be isolated from, for example and without limitation, the PKS genes encoding DEBS, the narbonolide PKS, and the FK-520 PKS. Each of the resulting PKS enzymes produces a polyketide compound that comprises a functional group at the C-11 position that can be further derivatized *in vitro* by standard chemical methodology to yield semi-synthetic epothilone derivatives of the invention.

20 The sixth module of the epothilone PKS includes a KS, an AT that binds methylmalonyl CoA, a DH, an ER, a KR, and an ACP. This module is encoded by a sequence within an ~14.5 kb HindIII-NsiI restriction fragment of cosmid pKOS35-70.1A2.

25 The recombinant DNA compounds of the invention that encode the sixth module of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the epothilone sixth module is inserted into a DNA compound that comprises the coding sequence for one or more modules of a heterologous PKS. The resulting protein encoded by the construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth module of the epothilone PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS when coexpressed with the other proteins comprising the PKS. In another embodiment, a DNA compound comprising a sequence that encodes the sixth module of

5 the epothilone PKS is inserted into a DNA compound that comprises the coding sequence
for modules 3, 4, and 5 of the epothilone PKS or a recombinant epothilone PKS that
produces an epothilone derivative and coexpressed with the other proteins of the
10 epothilone or epothilone derivative PKS to produce a PKS that makes epothilone or an
5 epothilone derivative in a host cell.

In another embodiment, a portion of the sixth module coding sequence is utilized
in conjunction with other PKS coding sequences to create a hybrid module. In this
15 embodiment, the invention provides, for example, either replacing the methylmalonyl CoA
specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific
10 AT; deleting any one, two, or all three of the ER, DH, and KR; and/or replacing any one,
two, or all three of the ER, DH, and KR with either a KR, a DH and KR, or a KR, DH, and
20 ER, including, optionally, to specify a different stereochemistry. In addition, the KS and/or
ACP can be replaced with another KS and/or ACP. In each of these replacements or
insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate
25 15 from a coding sequence for another module of the epothilone PKS, from a coding
sequence for a PKS that produces a polyketide other than epothilone, or from chemical
synthesis. The resulting heterologous sixth module coding sequence can be utilized in
conjunction with a coding sequence for a protein subunit of a PKS that makes epothilone,
30 an epothilone derivative, or another polyketide. If the PKS makes epothilone or an
20 epothilone derivative, the hybrid sixth module is typically expressed as a protein
comprising modules 3, 4, and 5 of the epothilone PKS or derivatives thereof.
Alternatively, the sixth module of the epothilone PKS can be deleted or replaced in its
35 entirety by a module from a heterologous PKS to produce a PKS for an epothilone
derivative.

25 Illustrative recombinant PKS genes of the invention include those in which the AT
40 domain encoding sequences for the sixth module of the epothilone PKS have been altered
or replaced to change the AT domain encoded thereby from a methylmalonyl specific AT
to a malonyl specific AT. Such malonyl specific AT domain encoding nucleic acids can be
45 isolated from, for example and without limitation, the PKS genes encoding the
30 narbonolide PKS, the rapamycin PKS, and the FK-520 PKS. When a recombinant *epoD*
gene of the invention encoding such a hybrid module 6 is coexpressed with the other
epothilone PKS genes, the recombinant PKS makes the 8-desmethyl epothilone
50 derivatives. This recombinant *epoD* gene derivative can also be coexpressed with

5 recombinant *epo* gene derivatives containing other alterations or can itself be further
altered to produce a PKS that makes the corresponding 8-desmethyl epothilone
derivatives. For example, one recombinant *epoD* gene provided by the invention also
10 includes module 4 coding sequences that encode an AT domain that binds only
5 methylmalonyl CoA. When incorporated into a PKS with the *epoA*, *epoB*, *epoC*, *epoE*,
and *epoF* genes, the recombinant *epoD* gene product leads to the production of the 8-
desmethyl derivatives of epothilones B and D.

15 Other illustrative recombinant *epoD* gene derivatives of the invention include those
in which the ER, DH, and KR domain encoding sequences for the sixth module of the
10 epothilone PKS have been replaced with those that encode (i) a KR and DH domain; (ii) a
KR domain; and (iii) an inactive KR domain. These recombinant *epoD* gene derivatives of
20 the invention, when coexpressed with the other epothilone PKS genes make the
corresponding (i) C-9 alkene, (ii) C-9 hydroxy, and (iii) C-9 keto epothilone derivatives.
These recombinant *epoD* gene derivatives can also be coexpressed with other recombinant
25 *epo* gene derivatives containing other alterations or can themselves be further altered to
produce a PKS that makes the corresponding C-9 epothilone derivatives. For example, one
recombinant *epoD* gene derivative provided by the invention also includes module 4
coding sequences that encode an AT domain that binds only methylmalonyl CoA. When
30 incorporated into a PKS with the *epoA*, *epoB*, *epoC*, *epoE*, and *epoF* genes, the
20 recombinant *epoD* gene product leads to the production of the C-9 derivatives of
epothilones B and D.

35 Functionally equivalent sixth modules can also be made by inactivation of one,
two, or all three of the ER, DH, and KR domains of the epothilone sixth module. The
preferred mode for altering such domains in any module is by replacement with the
25 complete set of desired domains taken from another module of the same or a heterologous
40 PKS coding sequence. Illustrative replacement domains for the substitutions described
above include but are not limited to the inactive KR domain from the rapamycin PKS
module 3 to form the ketone, the KR domain from the rapamycin PKS module 5 to form
the alcohol, and the KR and DH domains from the rapamycin PKS module 4 to form the
45 alkene. Other such inactive KR, active KR, and active KR and DH domain encoding
30 nucleic acids can be isolated from for example and without limitation the PKS genes
encoding DEBS, the narbonolide PKS, and the FK-520 PKS. Each of the resulting PKSs
50 produces a polyketide compound that comprises a functional group at the C-9 position that

5 can be further derivatized *in vitro* by standard chemical methodology to yield semi-synthetic epothilone derivatives of the invention.

10 The seventh module of the epothilone PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, and an ACP. This module is encoded by a sequence within an
5 ~8.7 kb BglII restriction fragment from cosmid pKOS35-70.4.

15 The recombinant DNA compounds of the invention that encode the seventh module of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. The seventh module of the epothilone PKS is contained in the gene product of the *epoE* gene, which also contains the eighth module.
10 The present invention provides the *epoE* gene in recombinant form, but also provides DNA compounds that encode the seventh module without coding sequences for the eighth module as well as DNA compounds that encode the eighth module without coding sequences for the seventh module. In one embodiment, a DNA compound comprising a sequence that encodes the epothilone seventh module is inserted into a DNA compound
20 that comprises the coding sequence for one or more modules of a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the seventh module of the epothilone PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel
30 PKS coding sequence that can be expressed in a host cell. Alternatively, the epothilone seventh module can be expressed as a discrete protein. In another embodiment, a DNA compound comprising a sequence that encodes the seventh module of the epothilone PKS is expressed to form a protein that, together with other proteins, constitutes the epothilone
35 PKS or a PKS that produces an epothilone derivative. In these embodiments, the seventh module is typically expressed as a protein comprising the eighth module of the epothilone
25 PKS or a derivative thereof and coexpressed with the *epoA*, *epoB*, *epoC*, *epoD*, and *epoF* genes or derivatives thereof to constitute the PKS.
40

45 In another embodiment, a portion or all of the seventh module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA
30 specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR; replacing the KR with a KR that specifies a different stereochemistry; and/or inserting a DH or a DH and an ER. In addition, the KS and/or
50 ACP can be replaced with another KS and/or ACP. In each of these replacements or
55

5 insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the epothilone PKS, from a coding sequence for a PKS that produces a polyketide other than epothilone, or from chemical
10 synthesis. The resulting heterologous seventh module coding sequence is utilized, optionally in conjunction with other coding sequences, to express a protein that together with other proteins constitutes a PKS that synthesizes epothilone, an epothilone derivative, or another polyketide. When used to prepare epothilone or an epothilone derivative, the
15 seventh module is typically expressed as a protein comprising the eighth module or derivative thereof and coexpressed with the *epoA*, *epoB*, *epoC*, *epoD*, and *epoF* genes or derivatives thereof to constitute the PKS. Alternatively, the coding sequences for the
20 seventh module in the *epoE* gene can be deleted or replaced by those for a heterologous module to prepare a recombinant *epoE* gene derivative that, together with the *epoA*, *epoB*, *epoC*, *epoD*, and *epoF* genes, can be expressed to make a PKS for an epothilone derivative.

25 15 Illustrative recombinant *epoE* gene derivatives of the invention include those in which the AT domain encoding sequences for the seventh module of the epothilone PKS have been altered or replaced to change the AT domain encoded thereby from a
30 methylmalonyl specific AT to a malonyl specific AT. Such malonyl specific AT domain encoding nucleic acids can be isolated from for example and without limitation the PKS genes encoding the narbonolide PKS, the rapamycin PKS, and the FK-520 PKS. When
20 coexpressed with the other epothilone PKS genes, *epoA*, *epoB*, *epoC*, *epoD*, and *epoF*, or derivatives thereof, a PKS for an epothilone derivative with a C-6 hydrogen, instead of a
35 C-6 methyl, is produced. Thus, if the genes contain no other alterations, the compounds produced are the 6-desmethyl epothilones.

40 25 The eighth module of the epothilone PKS includes a KS, an AT specific for methylmalonyl CoA, inactive KR and DH domains, a methyltransferase (MT) domain, and an ACP. This module is encoded by a sequence within an ~10 kb NotI restriction fragment of cosmid pKOS35-79.85.

45 30 The recombinant DNA compounds of the invention that encode the eighth module of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the epothilone eighth module is inserted into a DNA compound that
50 comprises the coding sequence for one or more modules of a heterologous PKS. The

5 resulting construct, in which the coding sequence for a module of the heterologous PKS is
either replaced by that for the eighth module of the epothilone PKS or the latter is merely
added to coding sequences for modules of the heterologous PKS, provides a novel PKS
10 coding sequence that is expressed with the other proteins constituting the PKS to provide a
5 novel PKS. Alternatively, the eighth module can be expressed as a discrete protein that
can associate with other PKS proteins to constitute a novel PKS. In another embodiment, a
DNA compound comprising a sequence that encodes the eighth module of the epothilone
15 PKS is coexpressed with the other proteins constituting the epothilone PKS or a PKS that
produces an epothilone derivative. In these embodiments, the eighth module is typically
20 expressed as a protein that also comprises the seventh module or a derivative thereof.

20 In another embodiment, a portion or all of the eighth module coding sequence is
utilized in conjunction with other PKS coding sequences to create a hybrid module. In this
embodiment, the invention provides, for example, either replacing the methylmalonyl CoA
specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific
25 AT; deleting the inactive KR and/or the inactive DH; replacing the inactive KR and/or DH
with an active KR and/or DH; and/or inserting an ER. In addition, the KS and/or ACP can
be replaced with another KS and/or ACP. In each of these replacements or insertions, the
heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding
30 sequence for another module of the epothilone PKS, from a coding sequence for a PKS
20 that produces a polyketide other than epothilone, or from chemical synthesis. The resulting
heterologous eighth module coding sequence is expressed as a protein that is utilized in
conjunction with the other proteins that constitute a PKS that synthesizes epothilone, an
35 epothilone derivative, or another polyketide. When used to prepare epothilone or an
epothilone derivative, the heterologous or hybrid eighth module is typically expressed as a
25 recombinant *epoE* gene product that also contains the seventh module. Alternatively, the
coding sequences for the eighth module in the *epoE* gene can be deleted or replaced by
those for a heterologous module to prepare a recombinant *epoE* gene that, together with
40 the *epoA*, *epoB*, *epoC*, *epoD*, and *epoF* genes, can be expressed to make a PKS for an
epothilone derivative.

45 30 The eighth module of the epothilone PKS also comprises a methylation or
methyltransferase (MT) domain with an activity that methylates the epothilone precursor.
This function can be deleted to produce a recombinant *epoD* gene derivative of the
50 invention, which can be expressed with the other epothilone PKS genes or derivatives

5 thereof that makes an epothilone derivative that lacks one or both methyl groups,
depending on whether the AT domain of the eighth module has been changed to a malonyl
specific AT domain, at the corresponding C-4 position of the epothilone molecule. In
10 another important embodiment, the present invention provides recombinant DNA
5 compounds that encode a polypeptide with this methylation domain and activity and a
variety of recombinant PKS coding sequences that encode recombinant PKS enzymes that
incorporate this polypeptide. The availability of this MT domain and the coding sequences
15 therefor provides a significant number of new polyketides that differ from known
polyketides by the presence of at least an additional methyl group. The MT domain of the
10 invention can in effect be added to any PKS module to direct the methylation at the
corresponding location in the polyketide produced by the PKS. As but one illustrative
20 example, the present invention provides the recombinant nucleic acid compounds resulting
from inserting the coding sequence for this MT activity into a coding sequence for any one
or more of the six modules of the DEBS enzyme to produce a recombinant DEBS that
25 synthesizes a 6-deoxyerythronolide B derivative that comprises one or more additional
methyl groups at the C-2, C-4, C-6, C-8, C-10, and/or C-12 positions. In such constructs,
the MT domain can be inserted adjacent to the AT or the ACP.

30 The ninth module of the epothilone PKS includes a KS, an AT specific for malonyl
CoA, a KR, an inactive DH, and an ACP. This module is encoded by a sequence within an
20 ~14.7 HindIII-BglII kb restriction fragment of cosmid pKOS35-79.85.

35 The recombinant DNA compounds of the invention that encode the ninth module
of the epothilone PKS and the corresponding polypeptides encoded thereby are useful for
a variety of applications. The ninth module of the epothilone PKS is expressed as a
protein, the product of the *epoF* gene, that also contains the TE domain of the epothilone
25 PKS. The present invention provides the *epoF* gene in recombinant form, as well as DNA
40 compounds that encode the ninth module without the coding sequences for the TE domain
and DNA compounds that encode the TE domain without the coding sequences for the
ninth module. In one embodiment, a DNA compound comprising a sequence that encodes
the epothilone ninth module is inserted into a DNA compound that comprises the coding
45 sequence for one or more modules of a heterologous PKS. The resulting construct, in
30 which the coding sequence for a module of the heterologous PKS is either replaced by that
for the ninth module of the epothilone PKS or the latter is merely added to coding
50 sequences for the modules of the heterologous PKS, provides a novel PKS protein coding

5 sequence that when coexpressed with the other proteins constituting a PKS provides a novel PKS. The ninth module coding sequence can also be expressed as a discrete protein with or without an attached TE domain. In another embodiment, a DNA compound
10 comprising a sequence that encodes the ninth module of the epothilone PKS is expressed
5 as a protein together with other proteins to constitute an epothilone PKS or a PKS that produces an epothilone derivative. In these embodiments, the ninth module is typically expressed as a protein that also contains the TE domain of either the epothilone PKS or a
15 heterologous PKS.

In another embodiment, a portion or all of the ninth module coding sequence is
10 utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the malonyl CoA
20 specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxy malonyl CoA specific AT; deleting the KR; replacing the KR with a KR that specifies a different stereochemistry; and/or inserting a DH or a DH and an ER. In addition, the KS and/or
25 ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the epothilone PKS, from a coding
30 sequence for a PKS that produces a polyketide other than epothilone, or from chemical synthesis. The resulting heterologous ninth module coding sequence is coexpressed with
20 the other proteins constituting a PKS that synthesizes epothilone, an epothilone derivative, or another polyketide. Alternatively, the present invention provides a PKS for an
35 epothilone or epothilone derivative in which the ninth module has been replaced by a module from a heterologous PKS or has been deleted in its entirety. In the latter embodiment, the TE domain is expressed as a discrete protein or fused to the eighth
25 module.

40 The ninth module of the epothilone PKS is followed by a thioesterase domain. This domain is encoded in the ~14.7 kb HindIII-BglII restriction comprising the ninth module coding sequence. The present invention provides recombinant DNA compounds that
45 encode hybrid PKS enzymes in which the ninth module of the epothilone PKS is fused to
30 a heterologous thioesterase or one or more modules of a heterologous PKS are fused to the epothilone PKS thioesterase. Thus, for example, a thioesterase domain coding sequence from another PKS can be inserted at the end of the ninth module ACP coding sequence in
50 recombinant DNA compounds of the invention. Recombinant DNA compounds encoding

5 this thioesterase domain are therefore useful in constructing DNA compounds that encode a protein of the epothilone PKS, a PKS that produces an epothilone derivative, and a PKS that produces a polyketide other than epothilone or an epothilone derivative.

10 In one important embodiment, the present invention thus provides a hybrid PKS
5 and the corresponding recombinant DNA compounds that encode the proteins constituting those hybrid PKS enzymes. For purposes of the present invention a hybrid PKS is a recombinant PKS that comprises all or part of one or more modules, loading domain, and
15 thioesterase/cyclase domain of a first PKS and all or part of one or more modules, loading domain, and thioesterase/cyclase domain of a second PKS. In one preferred embodiment,
10 the first PKS is most but not all of the epothilone PKS, and the second PKS is only a portion or all of a non-epothilone PKS. An illustrative example of such a hybrid PKS
20 includes an epothilone PKS in which the natural loading domain has been replaced with a loading domain of another PKS. Another example of such a hybrid PKS is an epothilone PKS in which the AT domain of module four is replaced with an AT domain from a
25 15 heterologous PKS that binds only methylmalonyl CoA. In another preferred embodiment, the first PKS is most but not all of a non-epothilone PKS, and the second PKS is only a portion or all of the epothilone PKS. An illustrative example of such a hybrid PKS
30 includes an erythromycin PKS in which an AT specific for methylmalonyl CoA is replaced with an AT from the epothilone PKS specific for malonyl CoA. Another example
20 is an erythromycin PKS that includes the MT domain of the epothilone PKS.

Those of skill in the art will recognize that all or part of either the first or second
35 PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For example, only a small portion of an AT domain determines its specificity. See U.S. patent application Serial No. 09/346,860 and PCT patent application No. WO
25 US99/15047, each of which is incorporated herein by reference. The state of the art in
40 DNA synthesis allows the artisan to construct de novo DNA compounds of size sufficient to construct a useful portion of a PKS module or domain. For purposes of the present invention, such synthetic DNA compounds are deemed to be a portion of a PKS.

45 The following Table lists references describing illustrative PKS genes and
30 corresponding enzymes that can be utilized in the construction of the recombinant PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing polyketide tailoring and modification enzymes and
50

5 corresponding genes that can be employed to make the recombinant DNA compounds of the present invention.

Avermectin

10 5 U.S. Pat. No. 5,252,474 to Merck.
MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the
15 Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemadectin.

10 MacNeil *et al.*, 1992, Gene 115: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

20 Ikeda and Omura, 1997, Chem. Res. 97: 2599-2609, Avermectin biosynthesis.

Candicidin (FR008)

Hu *et al.*, 1994, Mol. Microbiol. 14: 163-172.

25 Erythromycin

PCT Pub. No. 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, Science 252:675-9.

30 Cortes *et al.*, 8 Nov. 1990, Nature 348:176-8, An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*.

35 Glycosylation Enzymes

PCT Pat. App. Pub. No. 97/23630 to Abbott.

FK-506

25 Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK-506, Eur. J. Biochem. 256: 528-534.

40 Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK-506, Eur. J. Biochem. 244: 74-80.

45 30 Methyltransferase

US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from *Streptomyces* MA6858. 31-O-desmethyl-FK-506 methyltransferase.

50

55

5

Motamedi *et al.*, 1996, Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK-506 and FK-520, J. Bacteriol. 178: 5243-5248.

FK-520

10

- 5 U.S. patent application Serial No. 09/154,083, filed 16 Sep. 1998.
U.S. patent application Serial No. 09/410,551, filed 1 Oct. 1999.
Nielsen *et al.*, 1991, Biochem. 30:5789-96.

15

Lovastatin

U.S. Pat. No. 5,744,350 to Merck.

10 **Narhomycin**

U.S. patent application Serial No. 60/107,093, filed 5 Nov. 1998.

20

Nemadectin

MacNeil *et al.*, 1993, *supra*.

Niddamycin

25

- 15 Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, J. Bacteriol. 179: 7515-7522.

Oleandomycin

30

- Swan *et al.*, 1994, Characterisation of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence, Mol. Gen. Genet. 242: 358-362.

20

35

U.S. patent application Serial No. 60/120,254, filed 16 Feb. 1999, Serial No. 09/_____, filed 28 Oct. 1999, claiming priority thereto by inventors S. Shah, M. Betlach, R. McDaniel, and L. Tang, attorney docket No. 30063-20029.00.

40

- 25 Olano *et al.*, 1998, Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring, Mol. Gen. Genet. 259(3): 299-308.

Picromycin

45

- 30 PCT patent application No. WO US99/11814, filed 28 May 1999.
U.S. patent application Serial No. 09/320,878, filed 27 May 1999.
U.S. patent application Serial No. 09/141,908, filed 28 Aug. 1998.

50

55

5

Xue *et al.*, 1998, Hydroxylation of macrolactones YC-17 and narbomycin is mediated by the pikC-encoded cytochrome P450 in *Streptomyces venezuelae*, Chemistry & Biology 5(11): 661-667.

10

Xue *et al.*, Oct. 1998, A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity, Proc. Natl. Acad. Sci. USA 95: 12111 12116.

Platenolide

15

EP Pat. App. Pub. No. 791,656 to Lilly.

Pradimicin

10

PCT Pat. Pub. No. WO 98/11230 to Bristol-Myers Squibb.

Rapamycin

20

Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, Proc. Natl. Acad. Sci. USA 92:7839-7843.

25

Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, Gene 169: 9-16.

Rifamycin

30

PCT Pat. Pub. No. WO 98/07868 to Novartis.

August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amiclatopsis mediterranei* S669, Chemistry & Biology, 5(2): 69-79.

35

***Sorangium* PKS**

U.S. patent application Serial No. 09/144,085, filed 31 Aug. 1998.

Soraphen

40

25 U.S. Pat. No. 5,716,849 to Novartis.

Schupp *et al.*, 1995, J. Bacteriology 177: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

45

30

Spiramycin

U.S. Pat. No. 5,098,837 to Lilly.

Activator Gene

50

U.S. Pat. No. 5,514,544 to Lilly.

55

5

Tylosin

U.S. Pat. No. 5,876,991 to Lilly.

EP Pub. No. 791,655 to Lilly.

10

5 Kuhstoss *et al.*, 1996, Gene 183:231-6., Production of a novel polyketide through the construction of a hybrid polyketide synthase.

Tailoring enzymes

15

Merson-Davies and Cundliffe, 1994, Mol. Microbiol. 13: 349-355. Analysis of five tylosin biosynthetic genes from the tylBA region of the *Streptomyces fradiae* genome.

20

10 As the above Table illustrates, there are a wide variety of PKS genes that serve as readily available sources of DNA and sequence information for use in constructing the hybrid PKS-encoding DNA compounds of the invention. Methods for constructing hybrid PKS-encoding DNA compounds are described without reference to the epothilone PKS in U.S. Patent Nos. 5,672,491 and 5,712,146 and U.S. patent application Serial Nos. 09/073,538, filed 6 May 1998, and 09/141,908, filed 28 Aug 1998, each of which is
15 incorporated herein by reference. Preferred PKS enzymes and coding sequences for the proteins which constitute them for purposes of isolating heterologous PKS domain coding sequences for constructing hybrid PKS enzymes of the invention are the soraphen PKS and the PKS described as a *Sorangium* PKS in the above table.

30

To summarize the functions of the genes cloned and sequenced in Example 1:

<u>Gene</u>	<u>Protein</u>	<u>Modules</u>	<u>Domains Present</u>
<i>epoA</i>	EpoA	Load	Ks ^y mAT ER ACP
35 <i>epoB</i>	EpoB	1	NRPS, condensation, heterocyclization, adenylation, thiolation, PCP
<i>epoC</i>	EpoC	2	KS mmAT DH KR ACP
<i>epoD</i>	EpoD	3	KS mAT KR ACP
40		4	KS mAT KR ACP
		5	KS mAT DH ER KR ACP
		6	KS mmAT DH ER KR ACP
<i>epoE</i>	EpoE	7	KS mmAT KR ACP
45		8	KS mmAT MT DH* KR* ACP
<i>epoF</i>	EpoF	9	KS mAT KR DH* ACP TE

20

50

55

NRPS – non-ribosomal peptide synthetase; KS – ketosynthase; mAT – malonyl CoA specifying acyltransferase; mmAT – methylmalonyl CoA specifying acyltransferase; DH – dehydratase; ER – enoylreductase; KR – ketoreductase; MT – methyltransferase; TE thioesterase; * – inactive domain.

- 5 The hybrid PKS-encoding DNA compounds of the invention can be and often are hybrids of more than two PKS genes. Even where only two genes are used, there are often two or more modules in the hybrid gene in which all or part of the module is derived from a second (or third) PKS gene. Illustrative examples of recombinant epothilone derivative PKS genes of the invention, which are identified by listing the specificities of the hybrid modules (the other modules having the same specificity as the epothilone PKS), include:
- 10 (a) module 4 with methylmalonyl specific AT (mm AT) and a KR and module 2 with a malonyl specific AT (m AT) and a KR;
- (b) module 4 with mM AT and a KR and module 3 with mM AT and a KR;
- (c) module 4 with mM AT and a KR and module 5 with mM AT and a ER, DH, and KR;
- 15 (d) module 4 with mM AT and a KR and module 5 with mM AT and a DH and KR;
- (e) module 4 with mM AT and a KR and module 5 with mM AT and a KR;
- 20 (f) module 4 with mM AT and a KR and module 5 with mM AT and an inactive KR;
- (g) module 4 with mM AT and a KR and module 6 with m AT and a ER, DH, and KR;
- 25 (h) module 4 with mM AT and a KR and module 6 with m AT and a DH and KR;
- (i) module 4 with mM AT and a KR and module 6 with m AT and a KR;
- 30 (j) module 4 with mM AT and a KR and module 6 with m AT and an inactive KR;
- (k) module 4 with mM AT and a KR and module 7 with m AT;
- (l) hybrids (c) through (f), except that module 5 has a m AT;
- (m) hybrids (g) through (j) except that module 6 has a mM AT; and
- 35 (n) hybrids (a) through (m) except that module 4 has a m AT.
- 40 The above list is illustrative only and should not be construed as limiting the invention, which includes other recombinant epothilone PKS genes and enzymes with not only two hybrid modules other than those shown but also with three or more hybrid modules.
- 45
- 50
- 55

5 Those of skill in the art will appreciate that a hybrid PKS of the invention includes
but is not limited to a PKS of any of the following types: (i) an epothilone or epothilone
derivative PKS that contains a module in which at least one of the domains is from a
heterologous module; (ii) an epothilone or epothilone derivative PKS that contains a
10 module from a heterologous PKS; (iii) an epothilone or epothilone derivative PKS that
5 contains a protein from a heterologous PKS; and (iv) combinations of the foregoing.

While an important embodiment of the present invention relates to hybrid PKS
15 genes, the present invention also provides recombinant epothilone PKS genes in which
there is no second PKS gene sequence present but which differ from the epothilone PKS
10 gene by one or more deletions. The deletions can encompass one or more modules and/or
can be limited to a partial deletion within one or more modules. When a deletion
20 encompasses an entire module other than the NRPS module, the resulting epothilone
derivative is at least two carbons shorter than the compound produced from the PKS from
which the deleted version was derived. The deletion can also encompass the NRPS
25 module and/or the loading domain, as noted above. When a deletion is within a module,
the deletion typically encompasses a KR, DH, or ER domain, or both DH and ER
domains, or both KR and DH domains, or all three KR, DH, and ER domains.

The catalytic properties of the domains and modules of the epothilone PKS and of
30 epothilone modification enzymes can also be altered by random or site specific
20 mutagenesis of the corresponding genes. A wide variety of mutagenizing agents and
methods are known in the art and are suitable for this purpose. The technique known as
35 DNA shuffling can also be employed. See, e.g., U.S. Patent Nos. 5,830,721; 5,811,238;
and 5,605,793; and references cited therein, each of which is incorporated herein by
reference.

25 Recombinant Manipulations

40 To construct a hybrid PKS or epothilone derivative PKS gene of the invention, or
simply to express unmodified epothilone biosynthetic genes, one can employ a technique,
described in PCT Pub. No. 98/27203 and U.S. patent application Serial Nos. 08/989,332,
45 filed 11 Dec. 1997, and 60/129,731, filed 16 April 1999, each of which is incorporated
30 herein by reference, in which the various genes of the PKS are divided into two or more,
often three, segments, and each segment is placed on a separate expression vector. In this
50 manner, the full complement of genes can be assembled and manipulated more readily for

5 heterologous expression, and each of the segments of the gene can be altered, and various
altered segments can be combined in a single host cell to provide a recombinant PKS of
the invention. This technique makes more efficient the construction of large libraries of
10 recombinant PKS genes, vectors for expressing those genes, and host cells comprising
5 those vectors. In this and other contexts, the genes encoding the desired PKS are not only
present on two or more vectors, but also can be ordered or arranged differently than in the
native producer organism from which the genes were derived. Various examples of this
15 technique as applied to the epothilone PKS are described in the Examples below. In one
embodiment, the *epoA*, *epoB*, *epoC*, and *epoD* genes are present on a first plasmid, and the
10 *epoE* and *epoF* and optionally either the *epoK* or the *epoK* and *epoL* genes are present on a
second (or third) plasmid.

20 Thus, in one important embodiment, the recombinant nucleic acid compounds of
the invention are expression vectors. As used herein, the term "expression vector" refers to
any nucleic acid that can be introduced into a host cell or cell-free transcription and
25 translation medium. An expression vector can be maintained stably or transiently in a cell,
15 whether as part of the chromosomal or other DNA in the cell or in any cellular
compartment, such as a replicating vector in the cytoplasm. An expression vector also
comprises a gene that serves to produce RNA that is translated into a polypeptide in the
30 cell or cell extract. Thus, the vector typically includes a promoter to enhance gene
20 expression but alternatively may serve to incorporate the relevant coding sequence under
the control of an endogenous promoter. Furthermore, expression vectors may typically
35 contain additional functional elements, such as resistance-conferring genes to act as
selectable markers and regulatory genes to enhance promoter activity.

The various components of an expression vector can vary widely, depending on the
25 intended use of the vector. In particular, the components depend on the host cell(s) in
40 which the vector will be used or is intended to function. Vector components for expression
and maintenance of vectors in *E. coli* are widely known and commercially available, as are
vector components for other commonly used organisms, such as yeast cells and
45 *Streptomyces* cells.

30 In one embodiment, the vectors of the invention are used to transform *Sorangium*
host cells to provide the recombinant *Sorangium* host cells of the invention. U.S. Pat. No.
5,686,295, incorporated herein by reference, describes a method for transforming
50 *Sorangium* host cells, although other methods may also be employed. *Sorangium* is a

5 convenient host for expressing epothilone derivatives of the invention in which the
recombinant PKS that produces such derivatives is expressed from a recombinant vector
in which the epothilone PKS gene promoter is positioned to drive expression of the
10 recombinant coding sequence. The epothilone PKS gene promoter is provided in
5 recombinant form by the present invention and is an important embodiment thereof. The
promoter is contained within an ~500 nucleotide sequence between the end of the
transposon sequences and the start site of the open reading frame of the *epoA* gene.
15 Optionally, one can include sequences from further upstream of this 500 bp region in the
promoter. Those of skill in the art will recognize that, if a *Sorangium* host that produces
10 epothilone is used as the host cell, the recombinant vector need drive expression of only a
portion of the PKS containing the altered sequences. Thus, such a vector may comprise
20 only a single altered epothilone PKS gene, with the remainder of the epothilone PKS
polypeptides provided by the genes in the host cell chromosomal DNA. If the host cell
naturally produces an epothilone, the epothilone derivative will thus be produced in a
25 15 mixture containing the naturally occurring epothilone(s).

Those of skill will also recognize that the recombinant DNA compounds of the
invention can be used to construct *Sorangium* host cells in which one or more genes
involved in epothilone biosynthesis have been rendered inactive. Thus, the invention
30 provides such *Sorangium* host cells, which may be preferred host cells for expressing
20 epothilone derivatives of the invention so that complex mixtures of epothilones are
avoided. Particularly preferred host cells of this type include those in which one or more
35 of any of the epothilone PKS gene ORFs has been disrupted, and/or those in which any or
more of the epothilone modification enzyme genes have been disrupted. Such host cells
are typically constructed by a process involving homologous recombination using a vector
25 that contains DNA homologous to the regions flanking the gene segment to be altered and
40 positioned so that the desired homologous double crossover recombination event desired
will occur.

Homologous recombination can thus be used to delete, disrupt, or alter a gene. In a
45 preferred illustrative embodiment, the present invention provides a recombinant
30 epothilone producing *Sorangium cellulosum* host cell in which the *epoK* gene has been
deleted or disrupted by homologous recombination using a recombinant DNA vector of
the invention. This host cell, unable to make the *epoK* epoxidase gene product is unable to
50 make epothilones A and B and so is a preferred source of epothilones C and D.

5 Homologous recombination can also be used to alter the specificity of a PKS
module by replacing coding sequences for the module or domain of a module to be altered
with those specifying a module or domain of the desired specificity. In another preferred
10 illustrative embodiment, the present invention provides a recombinant epothilone
5 producing *Sorangium cellulosum* host cell in which the coding sequence for the AT
domain of module 4 encoded by the *epoD* gene has been altered by homologous
recombination using a recombinant DNA vector of the invention to encode an AT domain
15 that binds only methylmalonyl CoA. This host cell, unable to make epothilones A, C, and
E is a preferred source of epothilones B, D, and F. The invention also provides
10 recombinant *Sorangium* host cells in which both alterations and deletions of epothilone
biosynthetic genes have been made. For example, the invention provides recombinant
20 *Sorangium cellulosum* host cells in which both of the foregoing alteration and deletion
have been made, producing a host cell that makes only epothilone D.

In similar fashion, those of skill in the art will appreciate the present invention
25 provides a wide variety of recombinant *Sorangium cellulosum* host cells that make less
complex mixtures of the epothilones than do the wild type producing cells as well as those
that make one or more epothilone derivatives. Such host cells include those that make only
epothilones A, C, and E; those that make only epothilones B, D, and F, those that make
30 only epothilone D; and those that make only epothilone C.

20 In another preferred embodiment, the present invention provides expression
vectors and recombinant *Myxococcus*, preferably *M. xanthus*, host cells containing those
expression vectors that express a recombinant epothilone PKS or a PKS for an epothilone
35 derivative. Presently, vectors that replicate extrachromosomally in *M. xanthus* are not
known. There are, however, a number of phage known to integrate into *M. xanthus*
25 chromosomal DNA, including Mx8, Mx9, Mx81, and Mx82. The integration and
attachment function of these phages can be placed on plasmids to create phage-based
expression vectors that integrate into the *M. xanthus* chromosomal DNA. Of these, phage
40 Mx9 and Mx8 are preferred for purposes of the present invention. Plasmid pPLH343,
described in Salmi *et al.*, Feb. 1998, Genetic determinants of immunity and integration of
45 temperate *Myxococcus xanthus* phage Mx8, *J. Bact.* 180(3): 614-621, is a plasmid that
30 replicates in *E. coli* and comprises the phage Mx8 genes that encode the attachment and
integration functions.

5

10

15

20

25

30

35

40

45

50

55

5 The promoter of the epothilone PKS gene functions in *Myxococcus xanthus* host cells. Thus, in one embodiment, the present invention provides a recombinant promoter for use in recombinant host cells derived from the promoter of the *Sorangium cellulosum* epothilone PKS gene. The promoter can be used to drive expression of one or more
10 5 epothilone PKS genes or another useful gene product in recombinant host cells. The invention also provides an epothilone PKS expression vector in which one or more of the epothilone PKS or epothilone modification enzyme genes are under the control of their own promoter. Another preferred promoter for use in *Myxococcus xanthus* host cells for purposes of expressing a recombinant PKS of the invention is the promoter of the pilA gene of *M. xanthus*. This promoter, as well as two *M. xanthus* strains that express high
15 10 levels of gene products from genes controlled by the pilA promoter, a pilA deletion strain and a pilS deletion strain, are described in Wu and Kaiser, Dec. 1997, Regulation of expression of the pilA gene in *Myxococcus xanthus*, J. Bact. 179(24):7748-7758, incorporated herein by reference. Optionally, the invention provides recombinant
20 15 *Myxococcus* host cells comprising both the pilA and pilS deletions. Another preferred promoter is the starvation dependent promoter of the sdcK gene.

Selectable markers for use in *Myxococcus xanthus* include kanamycin, tetracycline, chloramphenicol, zeocin, spectinomycin, and streptomycin resistance conferring genes.
30 20 The recombinant DNA expression vectors of the invention for use in *Myxococcus* typically include such a selectable marker and may further comprise the promoter derived from an epothilone PKS or epothilone modification enzyme gene.

35 The present invention provides preferred expression vectors for use in preparing the recombinant *Myxococcus xanthus* expression vectors and host cells of the invention. These vectors, designated plasmids pKOS35-82.1 and pKOS35-82.2 (Figure 3), are able to
40 25 replicate in *E. coli* host cells as well as integrate into the chromosomal DNA of *M. xanthus*. The vectors comprise the Mx8 attachment and integration genes as well as the pilA promoter with restriction enzyme recognition sites placed conveniently downstream. The two vectors differ from one another merely in the orientation of the pilA promoter on
45 30 the vector and can be readily modified to include the epothilone PKS and modification enzyme genes of the invention. The construction of the vectors is described in Example 2.

Especially preferred *Myxococcus* host cells of the invention are those that produce an epothilone or epothilone derivative or mixtures of epothilones or epothilone derivatives
50 at equal to or greater than 20 mg/L, more preferably at equal to or greater than 200 mg/L,

5 and most preferably at equal to or greater than 1 g/L. Especially preferred are *M. xanthus*
host cells that produce at these levels. *M. xanthus* host cells that can be employed for
purposes of the invention include the DZ1 (Campos *et al.*, 1978, J. Mol. Biol. 119: 167-
10 178, incorporated herein by reference), the TA-producing cell line ATCC 31046, DK1219
5 (Hodgkin and Kaiser, 1979, Mol. Gen. Genet. 171: 177-191, incorporated herein by
reference), and the DK1622 cell lines (Kaiser, 1979, Proc. Natl. Acad. Sci. USA 76: 5952-
5956, incorporated herein by reference).

15 In another preferred embodiment, the present invention provides expression
vectors and recombinant *Pseudomonas fluorescens* host cells that contain those expression
10 vectors and express a recombinant PKS of the invention. A plasmid for use in constructing
the *P. fluorescens* expression vectors and host cells of the invention is plasmid pRSF1010,
20 which replicates in *E. coli* and *P. fluorescens* host cells (see Scholz *et al.*, 1989, Gene
75:271-8, incorporated herein by reference). Low copy number replicons and vectors can
also be used. As noted above, the invention also provides the promoter of the *Sorangium*
25 15 *cellulosum* epothilone PKS and epothilone modification enzyme genes in recombinant
form. The promoter can be used to drive expression of an epothilone PKS gene or other
gene in *P. fluorescens* host cells. Also, the promoter of the soraphen PKS genes can be
used in any host cell in which a *Sorangium* promoter functions. Thus, in one embodiment,
30 the present invention provides an epothilone PKS expression vector for use in *P.*
20 *fluorescens* host cells.

35 In another preferred embodiment, the expression vectors of the invention are used
to construct recombinant *Streptomyces* host cells that express a recombinant PKS of the
invention. *Streptomyces* host cells useful in accordance with the invention include
S. coelicolor, *S. lividans*, *S. venezuelae*, *S. ambofaciens*, *S. fradiae*, and the like. Preferred
25 *Streptomyces* host cell/vector combinations of the invention include *S. coelicolor* CH999
40 and *S. lividans* K4-114 and K4-155 host cells, which do not produce actinorhodin, and
expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. Patent
No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and
45 09/181,833, filed 28 Oct. 1998. Especially preferred *Streptomyces* host cells of the
30 invention are those that produce an epothilone or epothilone derivative or mixtures of
epothilones or epothilone derivatives at equal to or greater than 20 mg/L, more preferably
at equal to or greater than 200 mg/L, and most preferably at equal to or greater than 1 g/L.
50 Especially preferred are *S. coelicolor* and *S. lividans* host cells that produce at these levels.

5 Also, species of the closely related genus *Saccharopolyspora* can be used to produce epothilones, including but not limited to *S. erythraea*.

10 The present invention provides a wide variety of expression vectors for use in *Streptomyces*. For replicating vectors, the origin of replication can be, for example and
5 without limitation, a low copy number replicon and vectors comprising the same, such as SCP2* (see Hopwood *et al.*, Genetic Manipulation of *Streptomyces*: A Laboratory manual (The John Innes Foundation, Norwich, U.K., 1985); Lydiat *et al.*, 1985, Gene 35: 223-
15 235; and Kieser and Melton, 1988, Gene 65: 83-91, each of which is incorporated herein by reference), SLP1.2 (Thompson *et al.*, 1982, Gene 20: 51-62, incorporated herein by
10 reference), and pSG5(ts) (Muth *et al.*, 1989, Mol. Gen. Genet. 219: 341-348, and Bierman *et al.*, 1992, Gene 116: 43-49, each of which is incorporated herein by reference), or a high
20 copy number replicon and vectors comprising the same, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, J. Gen. Microbiol. 129: 2703-2714; Vara *et al.*, 1989, J. Bacteriol. 171: 5782-5781; and Servin-Gonzalez, 1993, Plasmid 30: 131-140, each of which is
25 15 incorporated herein by reference). High copy number vectors are generally, however, not preferred for expression of large genes or multiple genes. For non-replicating and
integrating vectors and generally for any vector, it is useful to include at least an *E. coli*
30 origin of replication, such as from pUC, p1P, p11, and pBR. For phage based vectors, the phage phiC31 and its derivative KC515 can be employed (see Hopwood *et al.*, *supra*).
20 Also, plasmid pSET152, plasmid pSAM, plasmids pSE101 and pSE211, all of which integrate site-specifically in the chromosomal DNA of *S. lividans*, can be employed.

35 Typically, the expression vector will comprise one or more marker genes by which host cells containing the vector can be identified and/or selected. Useful antibiotic
resistance conferring genes for use in *Streptomyces* host cells include the ermE (confers
25 resistance to erythromycin and lincomycin), tsr (confers resistance to thiostrepton), aadA
40 (confers resistance to spectinomycin and streptomycin), aacC4 (confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), hyg (confers
resistance to hygromycin), and vph (confers resistance to viomycin) resistance conferring
45 genes.

30 The recombinant PKS gene on the vector will be under the control of a promoter, typically with an attendant ribosome binding site sequence. A preferred promoter is the
actI promoter and its attendant activator gene actII-ORF4, which is provided in the pRM1
50 and pRM5 expression vectors, *supra*. This promoter is activated in the stationary phase of

5 growth when secondary metabolites are normally synthesized. Other useful *Streptomyces*
promoters include without limitation those from the *crmE* gene and the *melC1* gene,
which act constitutively, and the *tipA* gene and the *merA* gene, which can be induced at
10 any growth stage. In addition, the T7 RNA polymerase system has been transferred to
5 *Streptomyces* and can be employed in the vectors and host cells of the invention. In this
system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of
the chromosome or in a vector under the control of the inducible *merA* promoter, and the
15 gene of interest is placed under the control of the T7 promoter. As noted above, one or
more activator genes can also be employed to enhance the activity of a promoter.
10 Activator genes in addition to the *actII-ORF4* gene discussed above include *dnrI*, *redD*,
and *ptpA* genes (see U.S. patent application Serial No. 09/181,833, *supra*), which can be
20 employed with their cognate promoters to drive expression of a recombinant gene of the
invention.

The present invention also provides recombinant expression vectors that drive
25 expression of the epothilone PKS and PKS enzymes that produce epothilone or epothilone
derivatives in plant cells. Such vectors are constructed in accordance with the teachings in
U.S. patent application Serial No. 09/114,083, filed 10 July 1998, and PCT patent
publication No. 99/02669, each of which is incorporated herein by reference. Plants and
30 plant cells expressing epothilone are disease resistant and able to resist fungal infection.
20 For improved production of an epothilone or epothilone derivative in any heterologous
host cells, including plant, *Myxococcus*, *Pseudomonas*, and *Streptomyces* host cells, one
can also transform the cell to express a heterologous phosphopantetheinyl transferase. See
35 U.S. patent application Serial No. 08/728,742, filed 11 Oct. 1996, and PCT patent
publication No. 97/13845, both of which are incorporated herein by reference.

25 In addition to providing recombinant expression vectors that encode the epothilone
or an epothilone derivative PKS, the present invention also provides, as discussed above,
40 DNA compounds that encode epothilone modification enzyme genes. As discussed above,
these gene products convert epothilones C and D to cpothilones A and B, and convert
epothilones A and B to epothilones E and F. The present invention also provides
45 recombinant expression vectors and host cells transformed with those vectors that express
30 any one or more of those genes and so produce the corresponding epothilone or epothilone
derivative. In one aspect, the present invention provides the *epoK* gene in recombinant
50

5 form and host cells that express the gene product thereof, which converts epothilones C and D to epothilones A and B, respectively.

10 In another important embodiment, and as noted above, the present invention provides vectors for disrupting the function of any one or more of the *epoL*, *epoK*, and any
5 of the ORFs associated with the epothilone PKS gene cluster in *Sorangium* cells. The invention also provides recombinant *Sorangium* host cells lacking (or containing inactivated forms of) any one or more of these genes. These cells can be used to produce
15 the corresponding epothilones and epothilone derivatives that result from the absence of any one or more of these genes.

10 The invention also provides non-*Sorangium* host cells that contain a recombinant epothilone PKS or a PKS for an epothilone derivative but do not contain (or contain non-functional forms of) any epothilone modification enzyme genes. These host cells of the
20 invention are expected produce epothilones G and H in the absence of a dehydratase activity capable of forming the C-12-C-13 alkene of epothilones C and D. This
25 dehydration reaction is believed to take place in the absence of the *epoL* gene product in *Streptomyces* host cells. The host cells produce epothilones C and D (or the corresponding epothilone C and D derivative) when the dehydratase activity is present and the P450
30 epoxidase and hydroxylase (that converts epothilones A and B to epothilones E and F, respectively) genes are absent. The host cells also produce epothilones A and B (or the
20 corresponding epothilone A and B derivatives) when the hydroxylase gene only is absent. Preferred for expression in these host cells is the recombinant epothilone PKS enzymes of
35 the invention that contain the hybrid module 4 with an AT specific for methylmalonyl CoA only, optionally in combination with one or more additional hybrid modules. Also preferred for expression in these host cells is the recombinant epothilone PKS enzymes of
25 the invention that contain the hybrid module 4 with an AT specific for malonyl CoA only, optionally in combination with one or more additional hybrid modules.
40

The recombinant host cells of the invention can also include other genes and corresponding gene products that enhance production of a desired epothilone or epothilone
45 derivative. As but one non-limiting example, the epothilone PKS proteins require
30 phosphopantetheinylation of the ACP domains of the loading domain and modules 2 through 9 as well as of the PCP domain of the NRPS. Phosphopantetheinylation is mediated by enzymes that are called phosphopantetheinyl transferases (PPTases). To
50 produce functional PKS enzyme in host cells that do not naturally express a PPTase able

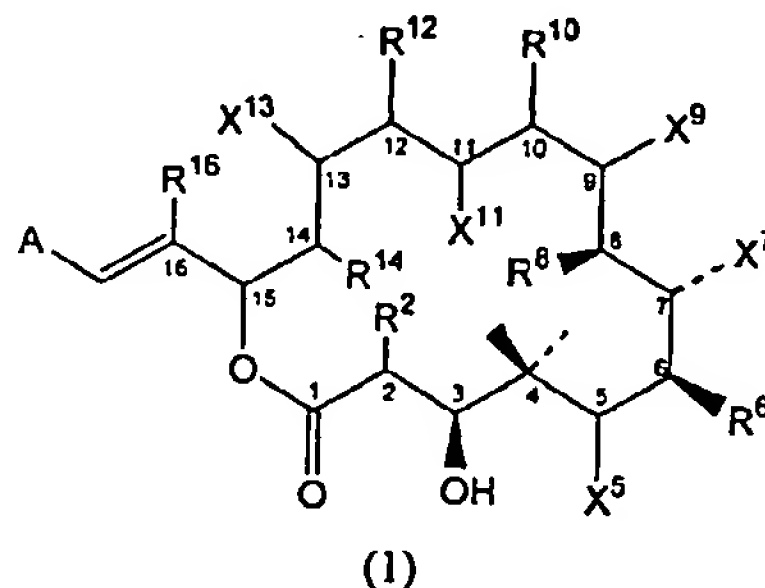
5 to act on the desired PKS enzyme or to increase amounts of functional PKS enzyme in
host cells in which the PPTase is rate-limiting, one can introduce a heterologous PPTase,
including but not limited to Sfp, as described in PCT Pat. Pub. Nos. 97/13845 and
10 98/27203, and U.S. patent application Serial Nos. 08/728,742, filed 11 Oct. 1996, and
5 08/989,332, each of which is incorporated herein by reference.

The host cells of the invention can be grown and fermented under conditions
known in the art for other purposes to produce the compounds of the invention. The
15 compounds of the invention can be isolated from the fermentation broths of these cultured
cells and purified by standard procedures. Fermentation conditions for producing the
10 compounds of the invention from *Sorangium* host cells can be based on the protocols
described in PCT patent publication Nos. 93/10121, 97/19086, 98/22461, and 99/42602,
20 each of which is incorporated herein by reference. The novel epothilone analogs of the
present invention, as well as the epothilones produced by the host cells of the invention,
can be derivatized and formulated as described in PCT patent publication Nos. 93/10121,
25 15 97/19086, 98/08849, 98/22461, 98/25929, 99/01124, 99/02514, 99/07692, 99/27890,
99/39694, 99/40047, 99/42602, 99/43653, 99/43320, 99/54319, 99/54319, and 99/54330,
and U.S. Patent No. 5,969,145, each of which is incorporated herein by reference.

30 Invention Compounds

20 Preferred compounds of the invention include the 14-methyl epothilone derivatives
(made by utilization of the hybrid module 3 of the invention that has an AT that binds
35 methylmalonyl CoA instead of malonyl CoA); the 8,9-dehydro epothilone derivatives
(made by utilization of the hybrid module 6 of the invention that has a DH and KR instead
of an ER, DH, and KR); the 10-methyl epothilone derivatives (made by utilization of the
25 hybrid module 5 of the invention that has an AT that binds methylmalonyl CoA instead of
40 malonyl CoA); the 9-hydroxy epothilone derivatives (made by utilization of the hybrid
module 6 of the invention that has a KR instead of an ER, DH, and KR); the 8-desmethyl-
14-methyl epothilone derivatives (made by utilization of the hybrid module 3 of the
45 invention that has an AT that binds methylmalonyl CoA instead of malonyl CoA and a
30 hybrid module 6 that binds malonyl CoA instead of methylmalonyl CoA); and the 8-
desmethyl-8,9-dehydro epothilone derivatives (made by utilization of the hybrid module 6
of the invention that has a DH and KR instead of an ER, DH, and KR and an AT that
50 specifies malonyl CoA instead of methylmalonyl CoA).

More generally, preferred epothilone derivative compounds of the invention are those that can be produced by altering the epothilone PKS genes as described herein and optionally by action of epothilone modification enzymes and/or by chemically modifying the resulting epothilones produced when those genes are expressed. Thus, the present invention provides compounds of the formula:



including the glycosylated forms thereof and stereoisomeric forms where the stereochemistry is not shown,

wherein A is a substituted or unsubstituted straight, branched chain or cyclic alkyl, alkenyl or alkynyl residue optionally containing 1-3 heteroatoms selected from O, S and N; or wherein A comprises a substituted or unsubstituted aromatic residue;

R^2 represents H, H, or H, lower alkyl, or lower alkyl, lower alkyl;

X^5 represents =O or a derivative thereof, or H, OH or H, NR₂ wherein R is H, or alkyl, or acyl or H, OCOR or H, OCONR₂ wherein R is H or alkyl, or is H, H;

R^6 represents H or lower alkyl, and the remaining substituent on the corresponding carbon is H;

X^7 represents OR, NR₂, wherein R is H, or alkyl or acyl or is OCOR, or OCONR₂ wherein R is H or alkyl or X^7 taken together with X^9 forms a carbonate or carbamate cycle, and wherein the remaining substituent on the corresponding carbon is H;

R^8 represents H or lower alkyl and the remaining substituent on the carbon is H;

X^9 represents =O or a derivative thereof, or is H, OR or H, NR₂, wherein R is H, or alkyl or acyl or is H, OCOR or H, OCONR₂ wherein R is H or alkyl, or represents H, H or wherein X^9 together with X^7 or with X^{11} can form a cyclic carbonate or carbamate;

R^{10} is H, H or H, lower alkyl, or lower alkyl, lower alkyl;

X^{11} is =O or a derivative thereof, or is H,OR, or H,NR₂ wherein R is H, or alkyl or acyl or is H,OCOR or H,OCONR₂ wherein R is H or alkyl, or is H,H or wherein X^{11} in combination with X^0 may form a cyclic carbonate or carbamate;

R^{12} is H,H, or H,lower alkyl, or lower alkyl,lower alkyl;

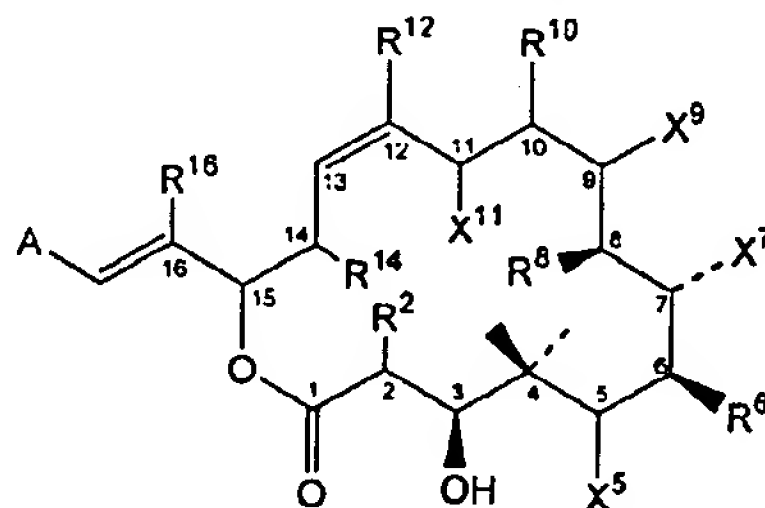
X^{13} is =O or a derivative thereof, or H,OR or H,NR₂ wherein R is H, alkyl or acyl or is H,OCOR or H,OCONR₂ wherein R is H or alkyl;

R^{14} is H,H, or H,lower alkyl, or lower alkyl,lower alkyl;

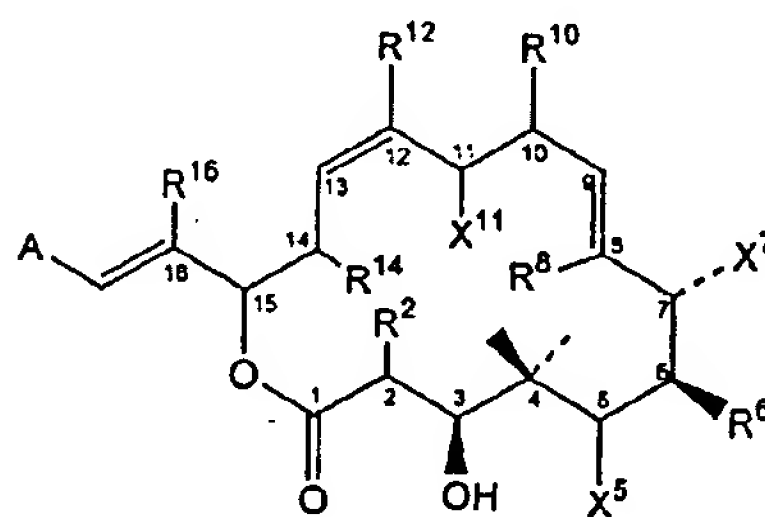
R^{16} is H or lower alkyl; and

wherein optionally II or another substituent may be removed from positions 12 and 13 and/or 8 and 9 to form a double bond, wherein said double bond may optionally be converted to an epoxide.

Particularly preferred are compounds of the formulas

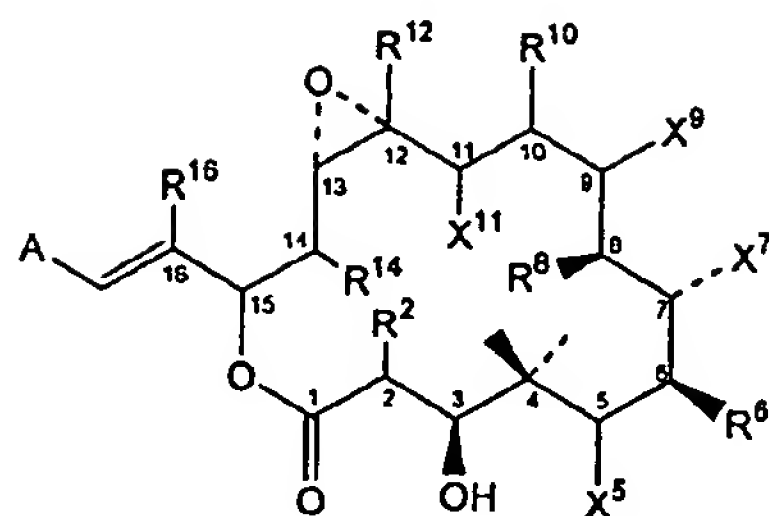


1(a),



1(b)

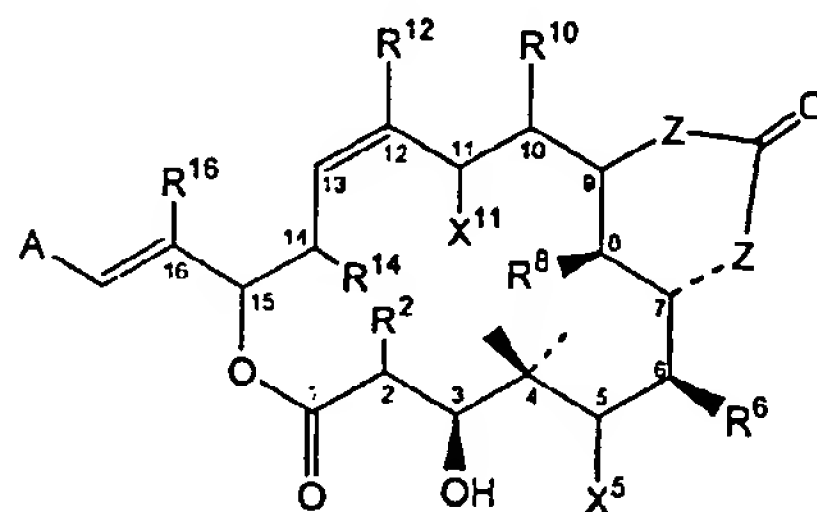
and



1(c)

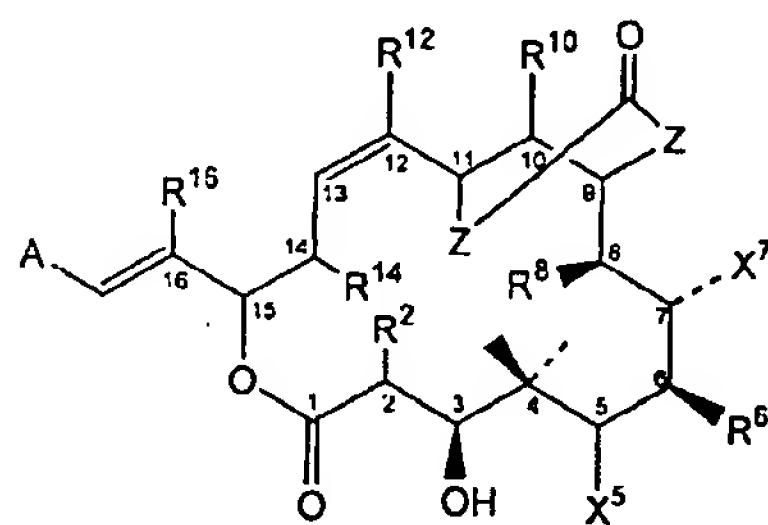
wherein the noted substituents are as defined above.

Especially preferred are compounds of the formulas



1(d)

and



1(e)

wherein both Z are O or one Z is N and the other Z is O, and the remaining substituents are as defined above.

As used herein, a substituent which "comprises an aromatic moiety" contains at least one aromatic ring, such as phenyl, pyridyl, pyrimidyl, thiophenyl, or thiazolyl. The substituent may also include fused aromatic residues such as naphthyl, indolyl, benzothiazolyl, and the like. The aromatic moiety may also be fused to a nonaromatic ring

5 and/or may be coupled to the remainder of the compound in which it is a substituent through a nonaromatic, for example, alkylene residue. The aromatic moiety may be substituted or unsubstituted as may the remainder of the substituent.

10 Preferred embodiments of A include the "R" groups shown in Figure 2.

5 As used herein, the term alkyl refers to a C₁-C₈ saturated, straight or branched chain hydrocarbon radical derived from a hydrocarbon moiety by removal of a single hydrogen atom. Alkenyl and alkynyl refer to the corresponding unsaturated forms. 15 Examples of alkyl include but are not limited to methyl, ethyl, propyl, isopropyl, n-butyl, tert-butyl, neopentyl, i-hexyl, n-heptyl, n-octyl. Lower alkyl (or alkenyl or alkynyl) refers 10 to a 1-4C radical. Methyl is preferred. Acyl refers to alkylCO, alkenylCO or alkynylCO.

20 The terms halo and halogen as used herein refer to an atom selected from fluorine, chlorine, bromine, and iodine. The term haloalkyl as used herein denotes an alkyl group to which one, two, or three halogen atoms are attached to any one carbon and includes without limitation chloromethyl, bromoethyl, trifluoromethyl, and the like.

25 15 The term heteroaryl as used herein refers to a cyclic aromatic radical having from five to ten ring atoms of which one ring atom is selected from S, O, and N; zero, one, or two ring atoms are additional heteroatoms independently selected from S, O, and N; and the remaining ring atoms are carbon, the radical being joined to the rest of the molecule 30 via any of the ring atoms, such as, for example, pyridyl, pyrazinyl, pyrimidinyl, pyrrolyl, pyrazolyl, imidazolyl, thiazolyl, oxazolyl, isoxazolyl, thiadiazolyl, oxadiazolyl, thiophenyl, furanyl, quinoliny, isoquinoliny, and the like. 20

35 The term heterocycle includes but is not limited to pyrrolidinyl, pyrazolinyl, pyrazolidinyl, imidazolinyl, imidazolidinyl, piperidinyl, piperazinyl, oxazolidinyl, isoxazolidinyl, morpholinyl, thiazolidinyl, isothiazolidinyl, and tetrahydrofuryl.

25 40 The term "substituted" as used herein refers to a group substituted by independent replacement of any of the hydrogen atoms thereon with, for example, Cl, Br, F, I, OH, CN, alkyl, alkoxy, alkoxy substituted with aryl, haloalkyl, alkylthio, amino, alkylamino, dialkylamino, mercapto, nitro, carboxaldehyde, carboxy, alkoxycarbonyl, or carboxamide. 45 Any one substituent may be an aryl, heteroaryl, or heterocycloalkyl group.

30 50 It will apparent that the nature of the substituents at positions 2, 4, 6, 8, 10, 12, 14 and 16 in formula (1) is determined at least initially by the specificity of the AT catalytic domain of modules 9, 8, 7, 6, 5, 4, 3 and 2, respectively. Because AT domains that accept malonyl CoA, methylmalonyl CoA, ethylmalonyl CoA (and in general, lower alkyl

malonyl CoA), as well as hydroxymalonyl CoA, are available, one of the substituents at these positions may be H, and the other may be H, lower alkyl, especially methyl and ethyl, or OH. Further reaction at these positions, e.g., a methyl transferase reaction such as that catalyzed by module 8 of the epothilone PKS, may be used to replace H at these positions as well. Further, an H,OH embodiment may be oxidized to =O or, with the adjacent ring C, be dehydrated to form a π -bond. Both OH and =O are readily derivatized as further described below.

Thus, a wide variety of embodiments of R^2 , R^6 , R^8 , R^{10} , R^{12} , R^{14} and R^{16} is synthetically available. The restrictions set forth with regard to embodiments of these substituents set forth in the definitions with respect to Formula (1) above reflect the information described in the SAR description in Example 8 below.

Similarly, β -carbonyl modifications (or absence of modification) can readily be controlled by modifying the epothilone PKS gene cluster to include the appropriate sequences in the corresponding positions of the epothilone gene cluster which will or will not contain active KR, DH and/or ER domains. Thus, the embodiments of X^5 , X^7 , X^9 , X^{11} and X^{13} synthetically available are numerous, including the formation of π -bonds with the adjacent ring positions.

Positions occupied by OH are readily converted to ethers or esters by means well known in the art; protection of OH at positions not to be derivatized may be required. Further, a hydroxyl may be converted to a leaving group, such as a tosylate, and replaced by an amino or halo substituent. A wide variety of "hydroxyl derivatives" such as those discussed above is known in the art.

Similarly, ring positions which contain oxo groups may be converted to "carbonyl derivatives" such as oximes, ketals, and the like. Initial reaction products with the oxo moieties may be further reacted to obtain more complex derivatives. As described in Example 8, such derivatives may ultimately result in a cyclic substituent linking two ring positions.

The enzymes useful in modification of the polyketide initially synthesized, such as transmethylnases, dehydratases, oxidases, glycosylation enzymes and the like, can be supplied endogenously by a host cell when the polyketide is synthesized intracellularly, by modifying a host to contain the recombinant materials for the production of these modifying enzymes, or can be supplied in a cell-free system, either in purified forms or as

5 relatively crude extracts. Thus, for example, the epoxidation of the π -bond at position 12-13 may be effected using the protein product of the *epoK* gene directly *in vitro*.

10 The nature of A is most conveniently controlled by employing an epothilone PKS which comprises an inactivated module 1 NRPS (using a module 2 substrate) or a KS2
5 knockout (using a module 3 substrate) as described in Example 6, hereinbelow. Limited variation can be obtained by altering the AT catalytic specificity of the loading module; further variation is accomplished by replacing the NRPS of module 1 with an NRPS of
15 different specificity or with a conventional PKS module. However, at present, variants are more readily prepared by feeding the synthetic module 2 substrate precursors and module
10 3 substrate precursors to the appropriately altered epothilone PKS as described in Example 6.
20

Pharmaceutical Compositions

25 The compounds can be readily formulated to provide the pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be
15 used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation will contain one or more of the compounds of the invention as an active ingredient in admixture with an organic or inorganic carrier or excipient
30 suitable for external, enteral, or parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, pessaries, solutions, emulsions, suspensions,
20 and any other form suitable for use.
35

The carriers which can be used include water, glucose, lactose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica,
25 potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquified form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl methylcellulose essentially as described in U.S. Patent
40 No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as
45 described in EPO patent publication No. 428,169, incorporated herein by reference.
30

Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, Transplantation Proceedings XIX, Supp. 6: 17-22, incorporated herein by reference.
50 Dosage forms for external application may be prepared essentially as described in EPO
55

5 patent publication No. 423,714, incorporated herein by reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

10 For the treatment of conditions and diseases caused by infection, immune system disorder (or to suppress immune function), or cancer, a compound of the invention may be administered orally, topically, parenterally, by inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, 15 adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intrathecal, intramuscular, and intrasternal injection or 20 infusion techniques.

20 Dosage levels of the compounds of the present invention are of the order from about 0.01 mg to about 100 mg per kilogram of body weight per day, preferably from about 0.1 mg to about 50 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 25 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the present invention may be administered on an intermittent basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

30 The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for oral administration to 35 humans may contain from 0.5 mg to 5 gm of active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds 40 of the invention may be formulated within the range of, for example, 0.00001% to 60% by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight.

45 It will be understood, however, that the specific dose level for any particular patient will depend on a variety of factors. These factors include the activity of the specific 50 compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

5 A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the present invention and shall not be construed as being a limitation on the scope of the invention or claims.

10
5 Example 1

DNA Sequencing of Cosmid Clones and Subclones Thereof

15 The epothilone producing strain, *Sorangium cellulosum* SMP44, was grown on a cellulose-containing medium, see Bollag *et al.*, 1995, Cancer Research 55: 2325-2333, incorporated herein by reference, and epothilone production was confirmed by LC/MS
20 analysis of the culture supernatant. Total DNA was prepared from this strain using the procedure described by Jaoua *et al.*, 1992, Plasmid 28: 157-165, incorporated herein by reference. To prepare a cosmid library, *S. cellulosum* genomic DNA was partially digested with Sau3AI and ligated with BamHI-digested pSupercos (Stratagene). The DNA was
25 packaged in lambda phage as recommended by the manufacturer and the mixture then used to infect *E. coli* XL1-Blue MR cells. This procedure yielded approximately 3,000 isolated colonies on LB-ampicillin plates. Because the size of the *S. cellulosum* genome is estimated to be circa 10^7 nucleotides, the DNA inserts present among 3000 colonies would correspond to circa 10 *S. cellulosum* genomes.

30 To screen the library, two segments of KS domains were used to design
20 oligonucleotide primers for a PCR with *Sorangium cellulosum* genomic DNA as template. The fragment generated was then used as a probe to screen the library. This approach was
35 chosen, because it was found, from the examination of over a dozen PKS genes, that KS domains are the most highly conserved (at the amino acid level) of all the PKS domains examined. Therefore, it was expected that the probes produced would detect not only the
25 epothilone PKS genes but also other PKS gene clusters represented in the library. The two degenerate oligonucleotides synthesized using conserved regions within the ketosynthase (KS) domains compiled from the DEBS and soraphen PKS gene sequences were (standard nomenclature for degenerate positions is used): CTSGTSKCSSTBCACCTSGCSTGC and
40 TGAYRTGSGCGTTSGTSCCGSWGA. A single band of ~750 bp, corresponding to the
45 predicted size, was seen in an agarose gel after PCR employing the oligos as primers and *S. cellulosum* SMP44 genomic DNA as template. The fragment was removed from the gel and cloned in the HincII site of pUC118 (which is a derivative of pUC18 with an insert
50 sequence for making single stranded DNA). After transformation of *E. coli*, plasmid DNA

5 from ten independent clones was isolated and sequenced. The analysis revealed nine
unique sequences that each corresponded to a common segment of KS domains in PKS
genes. Of the nine, three were identical to a polyketide synthase gene cluster previously
10 isolated from this organism and determined not to belong to the epothilone gene cluster
5 from the analysis of the modules. The remaining six KS fragments were excised from the
vector, pooled, end-labeled with ^{32}P and used as probe in hybridizations with the colonies
containing the cosmid library under high stringency conditions.

15 The screen identified 15 cosmids that hybridized to the pooled KS probes. DNA
was prepared from each cosmid, digested with NotI, separated on an agarose gel, and
10 transferred to a nitrocellulose membrane for Southern hybridization using the pooled KS
fragments as probe. The results revealed that two of the cosmids did not contain KS-
20 hybridizing inserts, leaving 13 cosmids to analyze further. The blot was stripped of the
label and re-probed, under less stringent conditions, with labeled DNA containing the
sequence corresponding to the enoylreductase domain from module four of the DEBS
25 gene cluster. Because it was anticipated that the epothilone PKS gene cluster would
encode two consecutive modules that contain an ER domain, and because not all PKS
gene clusters have ER domain-containing modules, hybridization with the ER probe was
30 predicted to identify cosmids containing insert DNA from the epothilone PKS gene
cluster. Two cosmids were found to hybridize strongly to the ER probe, one hybridized
20 moderately, and a final cosmid hybridized weakly. Analysis of the restriction pattern of
the NotI fragments indicated that the two cosmids that hybridized strongly with the ER
35 probe overlapped one another. The nucleotide sequence was also obtained from the ends
of each of the 13 cosmids using the T7 and T3 primer binding sites. All contained
sequences that showed homology to PKS genes. Sequence from one of the cosmids that
25 hybridized strongly to the ER probe showed homology to NRPSs and, in particular, to the
adenylation domain of an NRPS. Because it was anticipated that the thiazole moiety of
40 epothilone might be derived from the formation of an amide bond between an acetate and
cysteine molecule (with a subsequent cyclization step), the presence of an NRPS domain
45 in a cosmid that also contained ER domain(s) supported the prediction that this cosmid
30 might contain all or part of the epothilone PKS gene cluster.

Preliminary restriction analysis of the 12 remaining cosmids suggested that three
might overlap with the cosmid of interest. To verify this, oligonucleotides were
50 synthesized for each end of the four cosmids (determined from the end sequencing

described above) and used as primer sets in PCRs with each of the four cosmid DNAs. Overlap would be indicated by the appearance of a band from a non-cognate primer-template reaction. The results of this experiment verified that two of the cosmids overlapped with the cosmid containing the NRPS. Restriction mapping of the three cosmids revealed that the cosmids did, in fact, overlap. Furthermore, because PKS sequences extended to the end of the insert in the last overlapping fragment, based on the assumption that the NRPS would map to the 5'-end of the cluster, the results also indicated that the 3' end of the gene cluster had not been isolated among the clones identified.

To isolate the remaining segment of the epothilone biosynthesis genes, a PCR fragment was generated from the cosmid containing the most 3'-terminal region of the putative gene cluster. This fragment was used as a probe to screen a newly prepared cosmid library of *Sorangium cellulosum* genomic DNA of again approximately 3000 colonies. Several hybridizing clones were identified; DNA was made from six of them. Analysis of NotI-digested fragments indicated that all contained overlapping regions. The cosmid containing the largest insert DNA that also had the shortest overlap with the cosmid used to make the probe was selected for further analysis.

Restriction maps were created for the four cosmids, as shown in Figure 1. Sequence obtained from one of the ends of cosmid pKOS35-70.8A3 showed no homology to PKS sequences or any associated modifying enzymes. Similarly, sequence from one end of cosmid pKOS35-79.85 also did not contain sequences corresponding to a PKS region. These findings supported the observation that the epothilone cluster was contained within the ~70 kb region encompassed by the four cosmid inserts.

To sequence the inserts in the cosmids, each of the NotI restriction fragments from the four cosmids was cloned into the NotI site of the commercially available pBluescript plasmid. Initial sequencing was performed on the ends of each of the clones. Analysis of the sequences allowed the prediction, before having the complete sequence, that there would be 10 modules in this PKS gene cluster, a loading domain plus 9 modules.

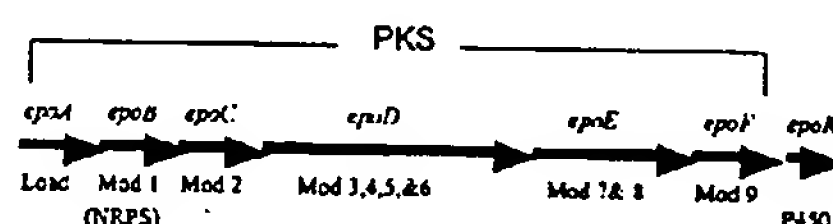
Sequence was obtained for the complete PKS as follows. Each of the 13 non-overlapping NotI fragments was isolated and subjected to partial HinPI digestion. Fragments of ~2 to 4 kb in length were removed from an agarose gel and cloned in the AccI site of pUC118. Sufficient clones from each library of the NotI fragments were sequenced to provide at least 4 -fold coverage of each. To sequence across each of the NotI sites, a set of oligos, one 5' and the other 3' to each NotI site, was made and used as

primers in PCR amplification of a fragment that contained each NotI site. Each fragment produced in this manner was cloned and sequenced.

The nucleotide sequence was determined for a linear segment corresponding to ~72 kb. Analysis revealed a PKS gene cluster with a loading domain and nine modules.

Downstream of the PKS sequence is an ORF, designated *epoK*, that shows strong homology to cytochrome P450 oxidase genes and encodes the epothilone epoxidase. The nucleotide sequence of 15 kb downstream of *epoK* has also been determined: a number of additional ORFs have been identified but an ORF that shows homology to any known dehydratase has not been identified. The *epoL* gene may encode a dehydratase activity, but this activity may instead be resident within the epothilone PKS or encoded by another gene.

The PKS genes are organized in 6 open reading frames. At the polypeptide level, the loading domain and modules 1, 2, and 9 appear on individual polypeptides; their corresponding genes are designated *epoA*, *epoB*, *epoC* and *epoF* respectively. Modules 3, 4, 5, and 6 are contained on a single polypeptide whose gene is designated *epoD*, and modules 7 and 8 are on another polypeptide whose gene is designated *epoE*. It is clear from the spacing between ORFs that *epoC*, *epoD*, *epoE* and *epoF* constitute an operon. The *epoA*, *epoB*, and *epoK* gene may be also part of the large operon, but there are spaces of approximately 100 bp between *epoB* and *epoC* and 115 bp between *epoF* and *epoK* which could contain a promoter. The present invention provides the intergenic sequences in recombinant form. At least one, but potentially more than one, promoter is used to express all of the epothilone genes. The epothilone PKS gene cluster is shown schematically below.



A detailed examination of the modules shows an organization and composition that is consistent with one able to be used for the biosynthesis of epothilone. The description that follows is at the polypeptide level. The sequence of the AT domain in the loading module and in modules 3, 4, 5, and 9 shows similarity to the consensus sequence for malonyl loading domains, consistent with the presence of an H side chain at C-14, C-12

5 (epothilones A and C), C-10, and C-2, respectively, as well as the loading region. The AT domains in modules 2, 6, 7, and 8 resemble the consensus sequence for methylmalonyl specifying AT domains, again consistent with the presence of methyl side chains at C-16, C-8, C-6, and C-4 respectively.

10 5 The loading module contains a KS domain in which the cysteine residue usually present at the active site is instead a tyrosine. This domain is designated as KS^y and serves as a decarboxylase, which is part of its normal function, but cannot function as a
15 condensing enzyme. Thus, the loading domain is expected to load malonyl CoA, move it to the ACP, and decarboxylate it to yield the acetyl residue required for condensation with
20 cysteine.

20 Module 1 is the non-ribosomal peptide synthetase that activates cysteine and catalyzes the condensation with acetate on the loading module. The sequence contains segments highly similar to ATP-binding and ATPase domains, required for activation of amino acids, a phosphopantotheinylation site, and an elongation domain. In database
25 15 searches, module 1 shows very high similarity to a number of previously identified peptide synthetases.

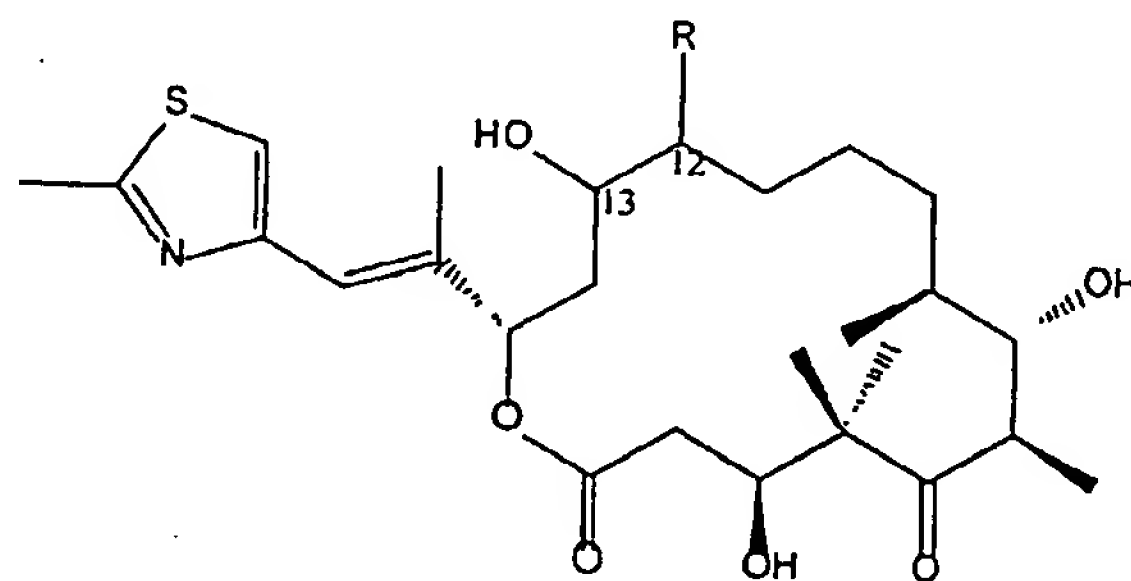
Module 2 determines the structure of epothilone at C-15 – C-17. The presence of the DH domain in module 2 yields the C-16-17 dehydro moiety in the molecule. The
30 domains in module 3 are consistent with the structure of epothilone at C-14 and C-15; the
20 OH that comes from the action of the KR is employed in the lactonization of the molecule.

Module 4 controls the structure at C-12 and C-13 where a double bond is found in epothilones C and D, consistent with the presence of a DH domain. Although the sequence
35 of the AT domain appears to resemble those that specify malonate loading, it can also load methylmalonate, thereby accounting in part for the mixture of epothilones found in the
25 fermentation broths of the naturally producing organisms.

40 A significant departure from the expected array of functions was found in module 4. This module was expected to contain a DH domain, thereby directing the synthesis of epothilones C and D as the products of the PKS. Rigorous analysis revealed that the space
45 between the AT and KR domains of module 4 was not large enough to accommodate a
30 functional DH domain. Thus, the extent of reduction at module 4 does not proceed beyond the ketoreduction of the beta-keto formed after the condensation directed by module 4. Because the C-12,13 unsaturation has been demonstrated (epothilones C and D), there
50 must be an additional dehydratase function that introduces the double bond, and this

function is believed to be in the PKS itself or resident in an ORF in the epothilone biosynthetic gene cluster.

Thus, the action of the dehydratase could occur either during the synthesis of the polyketide or after cyclization has taken place. In the former case, the compounds produced at the end of acyl chain growth would be epothilones C and D. If the C-12,13 dehydration were a post-polyketide event, the completed acyl chain would have a hydroxyl group at C-13, as shown below. The names epothilones G and H have been assigned to the 13-hydroxy compounds produced in the absence of or prior to the action of the dehydratase.



Epothilones G (R=H) and H (R=CH₃).

Modules 5 and 6 each have the full set of reduction domains (KR, DH and ER) to yield the methylene functions at C-11 and C-9. Modules 7 and 9 have KR domains to yield the hydroxyls at C-7 and C-3, and module 8 does not have a functional KR domain, consistent with the presence of the keto group at C-5. Module 8 also contains a methyltransferase (MT) domain that results in the presence of the geminal dimethyl function at C-4. Module 9 has a thioesterase domain that terminates polyketide synthesis and catalyzes ring closure. The genes, proteins, modules, and domains of the epothilone PKS are summarized in the Table hereinabove.

Inspection of the sequence has revealed translational coupling between *epoA* and *epoB* (loading domain and module 1) and between *epoC* and *epoD*. Very small gaps are seen between *epoD* and *epoE* and *epoE* and *epoF* but gaps exceeding 100 bp are found between *epoB* and *epoC* and *epoF* and *epoK*. These intergenic regions may contain promoters. Sequencing efforts have not revealed the presence of regulatory genes, and it is possible that epothilone synthesis is not regulated by operon specific regulation in *Sorangium cellulosum*.

5

The sequence of the epothilone PKS and flanking regions has been compiled into a single contig, as shown below.

10

15

20

25

30

35

40

45

50

55

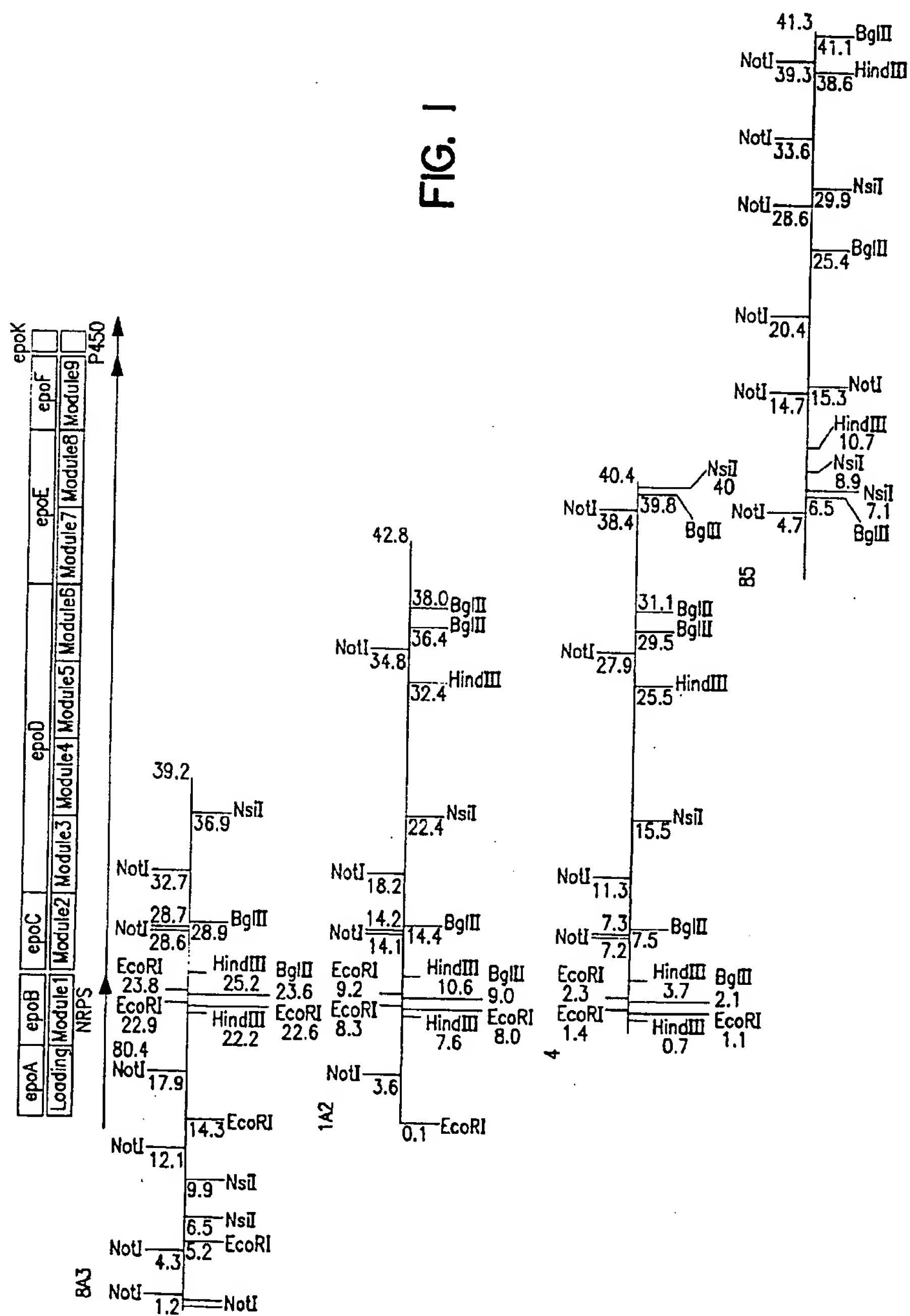
1 TCGTSCGCGG GCACGTCGAG GCGTTTGCCG ACTTCGGCGG CGTCCCGCGC GTGCTGCTCT
61 ACGACAACTT CAAGAACGCC GTCGTCGAGC GCCACGGCGA CGCGATCCGG TTCCACCCCA
121 CCTGCTGGC TCTGTCGGCG GATTACCGCT TCGAGCCGCG CCGCGTCCCG GTGCCCCGCG
181 GCAACGAGAA GGGCCGCGTC GAGCGCGCCA TCCGCTACGT CCGCGAGGGC TTCTTCGAGG
241 CCCGGGCTTA CGCCGACCTC GGAGACCTCA ACCGCCAAGC GACCGAGTGG ACCAGCTCCG
301 CCGCGCTCGA TCGCTCCTGG GTCGAGGACC GCGCCCGCAC CGTGCCTCAG GCCTTCGACG
361 ACGAGCGCAG CSTGCTGCTG CGACACCCTG ACACACCGTT TCCGGACCAC GAGCGCGTCC
421 AGGTCGAGGT CGGAAAGACC CCTACGCGC GCTTCGATCT CAACGACTAC TCGTCCCCC
481 ACGACCGGAC GCGCCGACCG CTGGTCGTCC TCGCCGACCT CAGTCAGGTA CGCATCGCCG
541 ACGGCAACCA GATCGTCGCG ACCACGTCCT GTTCGTGGGA CCGCGGCCAG CAGATCGAGC
601 AGCCCGAGCA CTTCCAGCGC CTGGTCGACC AGAAGCGCCG CGCCCGCGAG CACCGCGGCC
661 TTGATCGCCT CGCGCGCGCC GCGCGCAGCA GCCAGSCATT CCTGCGUATC GTCGCGGAGC
721 GCGGCGATAA CGTCGCGAGC GCGATCGCCC GGCTTCTGCA ACTGCTCGAC GCCGTGGGCG
781 CCGCCGAGCT CGAAGAGGCC CTGGTCGAGG TGCTTGAGCG CGACACCATC CACATCGGTG
841 CCGTCCGCCA GGTGATCGAC CGCGCCGCTT CCGAGCGCCA CCTGCCGCTT CCAGTCTCAA
901 TCCCCGTCAC CCGCGGCGAG CACGCCGCCC TCGTCTGTC GCGCGATTCC CTCACCACCT
961 ACGACGCCCT GAAGAAGGAC CCGACGCCAT GACCGACCTG ACGCCACCGG AGACCAAAGA
1021 CCGGCTCAAG AGCCTCGGCC TCTTCGGCCT GCTCGCCTGC TCGGAGCAGC TCGCCGACAA
1081 GCGCTGGCTT CGCGAGGTGC TCGCCATCGA GGAGCGCGAG CGCCACAAGC GCAGCCTCGA
1141 ACSCCGCCTG AAGAACTCCC GCGTCGCGCG CTTCAAGCCC ATGACCGACT TCGACTCGTC
1201 CTGGCCCAAG AAGATCGACC GCGAGGCGCT CGACGACCTC TACGATAGCC GCTACGCGGA
1261 CCTGCTCTTC GAGGTCTGTA CCGCTCGCTA CGACGCGCAG AAGCCGCTCT TGCTCAGCAC
1321 GAACAAGGCA TTCGCCGACT GGGGCCAGGT CTTCCCGCAC GCCCGGTCCG TCGTCACGCT
1381 CGTCGACCGG CTGCTGCAAC GCGCCGAGGT GATCGAGATC GAGGCCGAGA GCTACCGGCT
1441 GAAGGAAGCC AAGGAGCTCA ACGCCACCCG CACCAAGCAG CGCCGCACCA AGAAGCACTG
1501 AGCGGCATTT TCACCGGTGA ACTTCACCGA AATCCCGCGT GTTGCCGAGA TCATCTACAG
1561 GCGGATCGAG ACCGTGCTCA CCGCGTGGAC GACATGGCGC GGAAACGTCG TCGTAACGTC
1621 CCAGCAATGT CATGGGAATG GCGCCTTGAG GGGCTGGCCG GGGTCGACGA TATCGCGCGA
1681 TCTCCCCGTC AATTCCCGAG CGTAAAGAA AAATTTGTCA TAGATCGTAA GCTGTGCTAG
1741 TGATCTGCCT TACGTTACGT CTTCCGACCC TCGAGCGAAT TCTCTCGGAT AACTTTCAAG
1801 TTTTCTGAGG GGGCTTGCTC TCTGTTTCTT CAGGAAACCT GATCGGGACG AGCTAATTCC
1861 CATCCATTTT TTTGACACTC TGCTCAAAGG GATTAGACCG AGTGAGACAG TTCTTTTGCA
1921 GTGAGCGAAG AACCTGGGCG GCGGATCGAC GACGATCGAC GTCCCGGAGC GGGTCAGCCG
1981 CTGAGGATGT GCGCGTCTGT GCGGATCGTC CCATCGAGCG CGCAGCCGAA GATCCGATTG
2041 CGATCGTCCG AGCGGGCTGC CGTCTGCCCG GTGGCGTGAT CGATCTGAGC GGGTTCTGGA
2101 CGCTCCTCGA GGGCTCGCGC GACACCGTCG GGCAAGTCCC CCGCGAACGC TGGGATGCAG
2161 CAGCGTGGTT TGATCCCGAC CTCGATGCCC CGGGGAAGAC GCGCGTTACG CGCGCATCTT
2221 TCCTGAGCGA CGTAGCCTGC TTCGACGCTT CTTCTCTCGG CATCTCGCCT CGCGAAGCGC
2281 TGCGGATGGA CCTGACATAT CCGTCTGCTA CGGAAACGGG AGTGTTTCATC GGGATCGGCC
2341 CCGCGATCGC TCCATCGGCG CTCGTCGGTA CGGAAACGGG AGTGTTTCATC GGGATCGGCC
2401 CCGCGAATA TGAGGCGCGC CTGCGCGGAG CGACGGCGTC CGCAGAGATC GACGCTCATG
2461 GCGGGCTGGG CACCATGCCC AGCGTCGGAG CCGGCGGAAT CTCGTATGTC CTCGGGCTGC
2521 GAGGGCCGTC TGTGCGGTC GATACGGCCT ATTGCTCTTC GCTCGTGGCC STTCATCTGG
2581 CCTGTGAGAG CTTGCGCTCC GGGGAATGCT CCACGGCCCT GGCTGGTGGG STATCGCTGA
2641 TGTTGTGCGC GAGCACCTTC GTGTGGCTCT CGAAGACCCG CGCGCTGGCC ACGGACGCTC
2701 GCTGCAAGGC GTTTTCGGCG GAGGCGGATG GGTTCGGACG AGGCGAAGGG TGCGCCGTCG
2761 TGGTCCTCAA GCGGCTCAST GGAGCCCGCG CGGACGGCGA CCGGATATTG GCGGTGATTC
2821 GAGGATCCGC GATCAATCAC GACGGAGCGA GCAGCGGTCT GACCGTCCCG AACGGGAGCT
2881 CCGAAGAAAT CGTGCTGAAA CGCCCCCTGG CCGACGACAG CTGCGCCGCG TCTTCGGTGG
2941 GTTATGTCGA GGCACACGGC ACGGGCACGA CGCTTGGTGA CCGCATCGAA ATCCAAGCTC
3001 TGAATGCGGT ATACGGCCTC GGGCGAGACG TCGCCACGCC GCTGCTGATC GGGTCGGTGA
3061 AGACCAACCT TGGCCATCCT GAGTATGCGT CCGGGATCAC TGGGCTGCTG AAGGTCTGCT
3121 TGTCCCTTCA GCACGGGCGA ATTCTGCGC ACCTCCACGC GCAGCGCTG AACCCCGGGA
3181 TCTCATGGGG TGATCTTCGG CTGACCGTCA CGCGCGCCCG GACACCGTGG CCGGACTGGA
3241 ATACGCCGCG ACGGGCGGGG GTGAGCTCGT TCGGCATGAG CCGGACCAAC GCGCACGTGG
3301 TGCTGGAAGA GCGCCGCGCG GCGACGTGCA CACCGCCGCG GCGGAGCGCG CCGGACGAGC

5 3361 TGCTGGTGCT GTCGGCAAGG ACCGCGGCAG CCTTGGATGC ACACGCGGCG CGGCTGCGCG
3421 ACCATCTGGA GACCTACCCT TCGCAGTGTC TGGGCGATGT GGCGTTCAGT CTGGCGACGA
3481 CGCGCAGCGC GATGGAGCAC CGGCTCGCGG TGGCGGCGAC GTCGAGCGAG GGGCTGCGGG
5 3541 CAGCCCTGGA CGCTGCGGCG CAGGGACAGA CGCCGCCCGG TGTGGTGCGC GGTATCGCCG
3601 ATTCCCTCACG CGGCAAGCTC GCCTTTCTCT TCACCGGACA GGGGCGSCAG ACGCTGGGCA
3661 TGGGCGGTGG GCTGTATGAT GTATGGCCCG CGTTCCGCGA GCGGTTGAC CTGTGCGTGA
10 3721 GGCTGTTCAA CCAGGAGCTC GACCGGCGCG TCCGCGAGGT GATGTGGGCC GAACCGGCCA
3781 GCGTCGACGC CGCGCTGCTC GACCAGACAG CCTTTACCCA GCCGCGCGTG TTCACCTTCG
10 3841 AGTATGCGCT CGCCGCGCTG TGGCGGTCGT GGGCGGTAGA GCCGAGTTG GTCGCTGGCC
3901 ATAGCATCGG TGAGCTGGTG GCTGCCTGCG TGGCGGGCGT CTTCTCGCTT GAGGACGCGG
3961 TGTTCCTSGT GGCTGCGCGC GGGCGCCTGA TGCAGGCGCT GCCGCGCGGC GGGGCGATGG
4021 TGTGATCGC GCGCGCGGAG GCCGATGTGG CTGCTGCGGT GCGCGCGCAC GCA3CGTCGG
15 4081 TGTGATCGC GCGCGCTAAC GGTCCGGACC AGGTGGTCAT CGCGGGCGCC GGGCAACCCG
4141 TGCATGCGAT CGCGCGGGCG ATGGCGCGCG GCGGGGCGCG AACCAAGGCG CTCCACGTCT
15 4201 CGCATGCGTT CCACTCACCG CTCATGGCCC CGATGCTGGA GCGGTTCCGG CGTGTGGCCG
4261 AGTCSGTGAG CTACCGGCGG CCGTCGATCG TCCTGGTCAG CAATCTGAGC GGAAGGCTG
4321 GCACAGACGA GGTGAGCTCG CCGGGCTAT TCGTGGTCAG CAATCTGAGC GGAAGGCTG
4381 TCGCGGATGG AGTGAAGGCG CTGCACGCGG CCGGTGCGGG CACCTTCGTC GAGGTGCGTC
20 4441 CGAAATCGAC GCTGCTCGGC CTGGTGCTTG CCTGCCTGCC GGACGCGCGG CCGGCGCTGC
4501 TCGCATCGTC GCGCGCTGGG CTGACGAGC CAGCGACCGT GCTCGAGGCG CTCGGCGGGC
20 4561 TCTGGGCGGT CGGTGGCGTG GTCTCTGGG CCGGCGCTCT CCCCTCAGGG GGGCGGCGGG
4621 TGCCGCTGCC CACGTACCCT TGGCAGCGCG AGCGCTACTG GATCGACACG AAAGCCGACG
4681 ACGCGGCGCG TGGCGACCGC CGTCTCCCG GAGCGGGTCA CGACGAGGTC GAGAAGGGGG
4741 GCGCGGTGCG CCGCGGCGAC CCGCGCAGCG CTCGCTCGA CCATCCGCG CCCGAGAGCG
25 4801 GACGCCGGGA GAAGGTCGAG GCCGCGGGCG ACCGTCCGTT CCGGCTCGAG ATCGATGAGC
4861 CAGGCGGTCT CSATCGCCTG GTGCTTCGGG TCACGAGCG GCGCGCCCCCT GGTCTTGCGG
25 4921 AGGTGAGAT CGCGCTCGAC GCGCGGGGGC TCAGCTTCAA TGATGTCCAG CTCGCGCTGG
4981 GCATGCTGCG CGACGACCTG CCGGGAAGAG CCAACCTCC GCTGCTGCTC GGAGGCGAST
5041 GCGCCGCGCG CATCGTCGCG GTGGGCGAGG CCGTGACGCG CCTTGTGGTG GGCCAACCSG
30 5101 TCATCGCCCT TTCGGCGGGA GCGTTTGCTA CCCACGTAC CACGTGCGCT GCGCTGGTSC
5161 TGCCCTCGGC TCAGGCGCTC TCGGCGACCG AGGCGCGCG CATGCCGTC GCGTACCTGA
5221 CGCATGCTA CGCGCTCGAC GGAATAGCCC GCCTTCAGCC GGGGAGCGG GTGCTGATCC
30 5281 ACGCGGCGAC CCGCGGCGTC GGTCTCGCG CGCTGCAGTG GCGCGACAC GTGGGAGCCG
5341 AGSTCCATGC GACGGCGGCG ACGCCCGAGA AGCGCGCTA CCTGGAGTCG CTGGGCGTGC
35 5401 GGTATGTGAG CGATTCCCGC TCGGACCGGT TCGTCCCGA CGTGCAGCG TGGACGGGCG
5461 GCGAGGGAGT AGACGTCTG CTCAACTCGC TTTCGGGCGA GCTGATCGAC AAGAGTTTCA
5521 ATCTCCTGCG ATCGACGCG CGGTTTGTGG AGCTCGGCAA GCGCGACTGT TACCGGATA
5581 ACCAGCTCGG GCTGCGCGCG TTCTGCGCA ATCTCTCCT CTGCTGGTG GATCTCCGGG
35 5641 GGATGATGCT CGAGCGGCG GCGCGGGTCC GTGCGCTCT CGAGGAGTC CTCGGCTGA
5701 TCGCGGCGAG CGTGTTCACC CCTCCCCCA TCGCGACGCT CCGGATCGCT CGTGTGCGCG
5761 ATGCGTTCCG GAGCATGGCG CAGGCGCAGC ATCTTGGGAA GCTCGTACTC ACGTGGGTG
5821 ACCCGGAGGT CCAGATCCGT ATTCCGACCC ACGCAGGCG CCGCCCGTCC ACCGGGATC
5881 GGGATCTGCT CGACAGGCTC GCGTCAGCTG CGCCGCGCG GCGCGCGCG GCGCTGGAGG
5941 CGTTCCCTCCG TACGCGAGGT TCGCAGGTGC TCGCGACGCC CGAAATCAAG GTCGGCGCGG
40 6001 AGGCGCTGTT CACCCGCTC GGCATGGAUT CGCTCATGGC CGTGGAGCTG CGCAATCGTA
6061 TCGAGGCGAG CCTCAAGCTG AAGCTGICGA CGACGTTCT GTCCACGTCC CCCAATATCG
6121 CCTTGTGAC CCAAAACCTG TTGGATGCTC TCGCCACAGC TCTCTCCTG GAGCGGGTCC
6181 CGGCGGAGAA CCTACGGGCA GCGGTGCAA GCGACTTCGT CTCTCGGGC GCAGATCAAG
50 6241 ACTGGGAAT CATTGCCCTA TGACGATCAA TCAGCTTCTG AACGAGCTCG AGCACCAGGG
6301 TGTCAAGCTG GCGGCCGATG GGGAGCGCCT CCAGATACAG GCGCCCAAGA ACGCCCTGAA
6361 CCGGAACCTG CTCGCTCGAA TCTCCGAGCA CAAPAGCAG ATCCTGACGA TGCTCCGTCA
45 6421 GAGACTCCCC GCAGAGTCCA TCGTGCCCGC CCCAGCCGAG CCGCACSTTC CGTTTCTCT
6461 CACAGACATC CAAGGATCCT ACTGGCTGGG TCGGACAGGA GCGTTTACGG TCCCAGCGG
6541 GATCCACGCC TATCGCGAAT ACGACTGTAC GGATCTCGAC GTGGCGAGGC TGAGCCGCGC
55 6601 CTTCGGAAA GTCGTGCGCG GGCACGACAT GCTTCGGGCC CACACGCTGC CCGACATGAT
6661 GCAGGTGATC GAGCCTAAG TCGACGCCGA CATCGAGATC ATCGATCTGC GCGGGCTCGA
6721 CCGGAGCACA CGGGAAGCGA GGCTCGTATC GTTGCAGAT GCGATGTGCG ACCGCATCTA
6781 TGACACCGAG CGCCCTCCGC TCTATCACGT CGTCCGCTT CCGTGGACG AGCAGCAAAC
50 6841 CCTCTCGTG CTCAGTATCG ATCTCATTA CGTTGACCTA GCGAGCCTGT CCATCATCTT

5 6901 CAAGGATTGG CTCAGCTTCT ACGAAGATCC CGAGACCTCT CTCCCTGTCC TGGAGCTCTC
6961 GTACCGCGAC TATGTGCTCG CGCTGGAGTC TCGCAAGAAG TCTGAGGCGC ATCAACGATC
7021 GATGGATTAC TGGAGCGGC GCGTCGCCGA GCTCCACCT CCGCCGATGC TTCCGATGAA
5 7081 GGCCGATCCA TCTACCCTGA GGGAGATCCG CTTCCGGCAAC ACGGAGCAAT GGCTGCCGT
7141 GGAATCCTGG AGTCGATTGA AGCAGCGTGT CCGGGAGCGC GGGCTGACCC CGACGGGCGT
7201 CATTCTCGCT GCATTTTCCG AGGTGATCGG GCGCTGGAGC GCGAGCCCC GGTTTACGCT
10 7261 CAACATAACG CTCITCAACC GGCTCCCCGT CCATCCCGC GTGAACGATA TCACCGGGGA
7321 CTTACGTCG ATGGTCTCTC TGGACATCGA CACCACTCGC GACAAGAGCT TCGAACAGCG
7381 CGCTAAGCGT ATTCAAGAGC AGCTGTGGGA AGCGATGGAT CACTGCGACG TAAGCGGTAT
10 7441 CGAGGTCCAG CGAGAGGCCG CCCGGGTCTT GGGGATCCAA CGAGGCCCAT TGTCCCCGT
7501 GGTGCTCACG AGCGCGCTCA ACCAGCAAGT CGTTGGTGTG ACCTCGCTGC AGAGGCTCGG
7561 CACTCCGGTG TACACCAGCA CGCAGACTCC TCAGCTGCTG CTGGATCATC AGCTCTACGA
15 7621 GCACGATGGG GACCTCGTCC TCGCCTCGGA CATCGTCGAC GGAGTGTTC CGCCCGACCT
7681 TCTGACGAC ATGCTCGAAG CGTACGTCGC TTTCTTCGG CCGCTCACTG AGGAACCATG
15 7741 GAGTGAACAG ATGCGCTGTT CGCTTCCGCC TGCCAGCTA GAAGCGCGGG CGAGCGCAAA
7801 CGAGACCAAC TCGCTGCTGA GCGAGCATAC GCTGCACGGC CTGTTCCCGG CGCGGGTCGA
7861 GCAGCTGCCCT ATGCAGCTCG CCGTGGTGTG GCGCGCAAG ACGCTCACGT ACGAAGAGCT
7921 TTCGCGCCGT TCGCGCGCAC TTGGCGCGCG GCTGCGCGAG CAGGGGGCAC GCCCGAACAC
20 7981 ATTSCTCGCG GTGGTGATGG AGAAAGGCTG GGAGCAGGTT GTGCGGTTT TCGCGGTGCT
20 8041 CGAGTCAGGC GCGGCTTACG TGCCGATCGA TGCCGACCTA CCGCGGAGC GTATCCACTA
8101 CCTCCTCGAT CATGGTGAGG TAAAGCTCGT GCTGACGAG CCATGGCTGG ATGCGCAACT
8161 GTCATGGCCG CCGGGGATCC AGCGGCTGCT CGTGAGCGAT GCGCGCTCG AAGGCGACGG
8221 CGACCAGCTT CCGATGATGC CCATTCAGAC ACCTTCGGAT CTCCCGTATG TCATCTACAC
25 8281 CTCGGGATCC ACAGGGTTGC CCAAGGGGGT GATGATCGAT CATCGGGGTG CCGTCAACAC
8341 CATCCTGGAC ATCAACGAGC GCTTCGAAAT AGGGCCCGGA GACAGAGTGC TGGCGCTCTC
25 8401 CTCGCTGACC TTCGATCTCT CGGTCTACGA TGTGTTCCGG ATCTTGCGG CGGGCGGTAC
8461 GATCGTGGTG CCGGACGCGT CCAAGCTGCG CGATCCGGCG CATTCGGCAG CGTTGATCGA
8521 ACGAGAGAAG GTGACGGTGT GGAAGCTCGT CCGGGCGCTG ATGCGGATGC TCGTCGAGCA
30 8581 TTCCGAGGGT CGCCCCGATT CGCTCGCTAG ETCTCTGCGG CTTTCGCTGC TGAGCGGCGA
8641 CTGGATCCCG GTGGGCTGCG CTGGCGAGCT CAGGGCCATC AGGGCCGGCG TCTCGGTGAT
8701 CAGCCTGGGC GGGGCCACCG AAGCGTCGAT CTGGTCCATC GGGTACCCCG TGAGGAACGT
8761 CGATCCATCG TGGGCGAGCA TCCCTACGG CCGTCCGCTG CGCAACCAGA CGTTCCACGT
30 8821 GCTCGATGAG GCGCTCGAAC CGCGCCCGGT CTGGTTCCG GGGCAACTCT ACATTGGCGG
8881 GGTCCGACTG GCACTGGGCT ACTGCCGCGA TGAAGAGAAG ACGCGCAACA GCTTCCTCGT
35 8941 GCACCCCGAG ACCGGGGAGC GCTTCACAA GACCGGCGAT CTGGGCGCT ACCTGCCCGA
9001 TGGAAACATC GAGTTCATCG GCGGGGAGGA CAACCAATC AAGCTTCGCG GATACCGCGT
9061 TGAGCTCGGG GAAATCGAGG AAACGCTCAA GTGCGATCCG AACGTACCG ACGCGGTGAT
35 9121 TGTGCCCGTC GGAACGAGC CGGCGAACAA GCTCCTTCTA GCCTATGTGG TCCCGGAAGG
9181 CACACACGCC CGCGCTGCCG AGCAGGACCG GAGCCTCAAG ACCGAGCGG TCGACGCGAG
40 9241 AGCACACGCC GCGAAGCGG ACGGATTGAG CGACGCGSAG AGGGTGCAG TCAAGCTCGC
9301 TCGACACGGA CTCCGGAGGG ATCTGACGG AAAGCCCGTC GTGATCTGA CCGGGCTGGT
9361 TCCGCGGGAG GCGGGGCTGG ACGTCTACGC GCGTCGCCGT AGCGTCCGAA CGTTCTCGA
9421 GGCCCCGATT CCATTTGTTT AATTCCGGCG ATTCCTGAGC TGCTGAGCA GCGTGGAGCC
45 9481 CGACGGCGCG GCCCTTCCCA AATTCCGTTA TCCATCGGCT GGCAGCACGT ACCCGGTGCA
9541 AACCTACGCG TACGCCAAT CCGGUCGCT CGAGGGCGTG GACGAGGGCT TCTATTATTA
40 9601 CCACCCGTTT GAGCACCGTT TGCTGAAGGT CTCCGATCAC GGGATCGAGC GCGGAGCGCA
9661 CGTTCGCGAA AACTTCGACG TGTTGATGA AGCGCGGTTT GGCCTCTGT TCGTGGGCG
9721 GATCGATGCC ATCGAGTCGC TGTATGGATC GTTGTCACGA GAATTCGCG TGTGGAGGC
50 9781 CGGATATATG GCGCAGCTCC TGATGGAGCA GCGGCTTCTC TGCAACATCG GCGTCTGTCC
9841 GGTGGGTCAA TTCGATTTG AACAGGTTCC GCGGTTCTC GACCTGCGGC ATTCCGACGT
9901 TTACGTGCAC GGCATGCTGG GCGGGCGGGT AGACCCGCGG CAGTTCAGG TCTGTACGCT
9961 CCGTCAGGAT TCCTCACCAG GCGCGGCCAC GACGCGCGGC GCCCTCCCG GCCGCGATCA
45 10021 GCACTTCGCC GATATCCTTC GCGACTTCTT GAGGACCAAA CTACCCGAGT ACATGGTGCC
55 10081 TACAGTCTTC GTGGAGCTCG ATGCGTTGCC GCTGACGTC AACGGCAAGG TCGATCGTAA
10141 GGCCCTGCGC GAGCGGAAGG ATACCTCGTC GCCGCGGCAT TCGGGGCACA CGGCSCCAG
10201 GCACGCCTTG GAGGAGATCC TCGTTGCGGT CCGTACGGAG GTGCTCGGGC TGGAGGTGGT
10261 TGGGCTCCAG CAGAGCTTCG TCGATCTTGG TCGGACATCG ATTCACATCG TTCGATCAG
10321 GAGTCTGTTG CAGAAGAGGC TGGATAGGSA GATCGCCATC ACCGAGTTGT TCCAGTACCC
50 10381 GAACCTCGGC TCGCTGGCGT CCGGTTTGCG CCGAGACTCG APAGATCTAG AGCAGCGGCC

5
10
15
20
25
30
35
40
45
50
55

10441 GAACATGCAG GACCGAGTGG AGGCTCGGCG CAAGGGCAGG AGACGTAGCT AAGAGCGCCG
10501 AACAAAACCA GGCCGAGCGG GCCAATGAAC CGCAAGCCCG CCTGCGTCAC CCTGGGACTC
10561 ATCTGATCTG ATCGCGGGTA CGCGTCGCGG GTGTGCGCGT TGAGCCGTGT TGCTCGAACG
10621 CTGAGGAACG GTGAGCTCAT GGAAGAACAA GAGTCC'CCG CTATCGCAGT CATCGGCATG
5 10681 TCGGECCTTT TTCCGGGGGC GCGGGATCTG GACGAATTCT GGAGGAACCT TCGAGACGGC
10741 ACGGAGGCCG TGCAGCGCTT CTCCGAGCAG GAGCTCGCGG CGTCCGGAGT CGACCCAGCG
10801 CTGGTGCTGG ACCCGAACTA CGTCCGGGCG GGCAGCGTGC TCGAAGATGT CGACCGGTTT
10861 GACGCTGCTT TCTTCGGCAT CAGCCCGCGC GAGGCAGAGC TCATGGATCC GCAGCACCAG
10921 ATCTTCATGG AATGCGCCTG GGAGGCGCTG GAGAAGCCCG GATACGACCC GACAGCCTAC
10 10981 GAGGCTCTTA TCGGCGTGTA CGCCGGCCCC AACATGAGCT CGTACTTGAC GTCGAACCTC
11041 CACGAGCACC CAGCGATGAT GCGGTGGCCC GGCTGGTTTC AGACGTTGAT CGGCAACGAC
11101 AAGGATTACC TCGCGACCCA CGTCTCCTAC AGGCTGAATC TGAGAGGGCC GAGCATCTCC
11161 GTTCAAACCT OCTGCTCTAC CTCGCTCGTG GCGGTTCACT TGGCGTGCAT GAGCTCTCTG
15 11221 GACCGCGAGT GCGACATGGC GCTGGCCGGC GGGATTACCG TCCGGATCCC CCATCGAGCC
11281 GGCTATGTAT ATGCTGAGGG GGGCATCTTC TCTCCGACG GCCATTGCCG GGCCTTCGAC
11341 GCCAAGGCGA ACGGCACGAT CATGGGCAAC GGCTGCGGGG TTGTCTCTCT GAAGCCGCTG
11401 GACCGGGCGC TCTCCGATGG TGATCCCGTC CGCGCGGTCA TCCTTGGGTC TGCCACAAAC
11461 AACGACGGAG CSAGGAAGAT CCGGTTCACT GCGCCCACTG AGGTGGGCA GCGCAAGCG
11521 ATCATGGAGG CACTGGCGCT GGCAGGGGTC GAGGCCCGGT CCATCCAATA CATCGAGACC
20 11581 CACGGGACCG GCACGCTGCT CGGACACGCC ATCGAGACGG CGGCGTTGCG GCGGGTGTTC
11641 GATCGCGACG CTTGACCCCG GAGGTCTTGC GCGATCGGCT CCGTGAAGAC CGGCATCGGA
11701 CACCTCGAAT CCGCGGCTGG CATCGCCGGT TTGATCAAGA CCGTCTTGGC GCTGGAGCAC
11761 CCGCAGCTGC CCCCCAGCCT GAACTTCGAG TCTCTAACC CATCGATCGA TTTCGCGAGC
11821 AGCCCGTCTT ACGTCAATAC CTCTCTTAAG GATTGGAATA CCGGCTCGAC TCCGCGGCGG
25 11881 GCCGGCGTCA GCTCGTTCGG GATCGGCGGC ACCAACGCCC ATGTCGTGCT GGAGGAAGCA
11941 CCGCGGCGCA AGCTTCCAGC CGCGGCGGCC GCGCGCTCTG CCGAGCTCTT CGTCTCTCG
12001 GCCAAGAGCG CAGCGGCGCT GGA'G'CCGCS CGGCAACGGC TACGAGATCA TCTGCAGGCG
12061 CACCAGGGGC TTTCGTTGGG CGACGTCGCC TTCAGCCTGG CGACGACGCG CAGTCCCATG
12121 GASCACCGGC TCGCGATGGC GGCACCGTCG CGCGAGGCGT TGCGAGAGGG GCTCGACGCA
30 12181 GCGGCGCGAG GCCAGACCCC GCGGGGCGCC GTCCGTGGCC GCTGCTCCCC AGGCAACGTG
12241 CCGAAGGTGG TCTTCGICTT TCCCGGCCAG GGCTCTCAGT GGGTCGGTAT GGGCCGTCAG
12301 CTCCTGGCTG AGGAACCCGT CT'CCACCGG GCGCTTTCCG CGTCCGACCS GGCCATCCAG
12361 GCCGAAGCTG GTTGGTCCCT GCTCGCCGAG CTCGCCCGCG ACGAAGGGTC GTCCCAGATC
35 12421 GAGCGCATCG ACGTGGTGCA GCGGTGCTG TTCGCGCTCG CCGTGGCATT TCGGGCGCTG
12481 TGGCGGTCTG GGGGTGTCCG GCGCGACGTC GTGATCGGCC ACAGCATGGG CGAGGTAGCC
12541 GCGCGCATG TGGCCGGGGC GCTGTGCTCG GAGGATECGG TGGCGATCAT CTGCCGGCGC
12601 AGCCGGCTGC TCCGGCGCAT CAGCGGTCAG GCGGAGATGG CCGTGACCGA GCTGTGCTG
12661 GCGGAGGCCG AGGCAGCGCT CCGAGGCTAC GAGGATCGGG TGAGCGTGGC CGTGAGCAAC
12721 AGCCCGCGCT CGACGGTGCT CTCGGGCGAG CCGGCAACGA TCGCGAGGT GCTGTCTCTC
35 12781 CTGAACGCGA AGGGGGTGT'CT CTGCGCTCGG GTGAGSTGG ATGTCGCCAG CCACAGCCCG
12841 CAGSTCGACC CGCTGCGCGA GGACCTCTTG GCAGCGCTGG GCGGGCTCCG GCGCGTGGC
12901 GCTGCGGTGC CGATGCGCTC GACGGTGACG GCGGCCATGG TAGCGGGCCC GGAGCTCGGA
12961 GCGAATTACT GGATGAACAA TCTCAGGCAG CCTGTGCGCT TCGCCGAGGT AGTCCAGGCG
40 13021 CAGCTCCAAG GCGGCCACGG TCTGTTCTGT GAGATGAGCC CGCATCCGAT CCTAACGACT
13081 TCGGTGAGG AGATGCGGCG CGCGGCCAG CCGGCGGGCG CAGCGGTGGG CTCGCTGCGG
45 13141 CGAGGCGAGG ACGAGCGCCC GCGGATGCTG GAGGCGCTGG GCGCGCTGTG GCGCGAGGGC
13201 TACCCTGTAC CCTGGGGGCG GCTGTTTCCC GCGGGGGGGC GCGCGCTACC GCTGCCGACC
13261 TA'CCCTGGC AGCGCGAGCG GTACTGGATC CAAGCGCCGG CCAAGAGCGC CGCGGGCGAT
50 13321 CGCGCGGGCG TCGGTGCGGG GTACTGGATC CCGGCGGGCG CTCCTCGGTG AAATGCAGAC CCTATCAACC
13381 CAGACGAGCA CGCGGCTGTG GGAGACGACG CTGATCTCA AGCAGCTGCC GTGGCTCGGC
13441 GACCACCGGG TGCAGGGAGC GGTCGTGTTT CCGGGCGCGG CGTACCTGGA GATGGCGATT
45 13501 TCGTCGGGGG CCGAGGCTTT GGGCGATGCG CCATTGCAGA TAACCGACGT GGTGCTCGCC
13561 GAGGCGCTGG CCTTCGCGGG CGACGCGGCG GTGTTGGTCC AGGTGGTGAC GACGGAGCAG
13621 CCGTCGGGAC GGCTGCAGTT CCAGATCGCG AGCCGGGCGC CCGGCGCTCG CCACGCGTCC
55 13681 TTCCGGGTCC ACGCTCGCGG CGCGTTGCTC CGAGTGGAGC GCACCGAGGT CCCGCTGGG
13741 CTTACGCTTT CCGCCGTGCG CGCACGGCTC CAGGCCAGCA TGCCCGCCGC GGCCACCTAC
13801 GCGGAGCTGA CCGAGATGGG GCTGCACTAC GCGCCCTGCC TCCAGGGGAT TGCTGAGCTA
13861 TGGCGCGGTG AGGGCGAGGC GCTGGGACGG GTACGCTGC CCGACGCGGC CGGCTCGGCA
50 13921 GCGGAGTATC GGTTCATCC TGCGCTGCTG GACGCGTGT TCCAGGTCGT CCGCAGCCTC



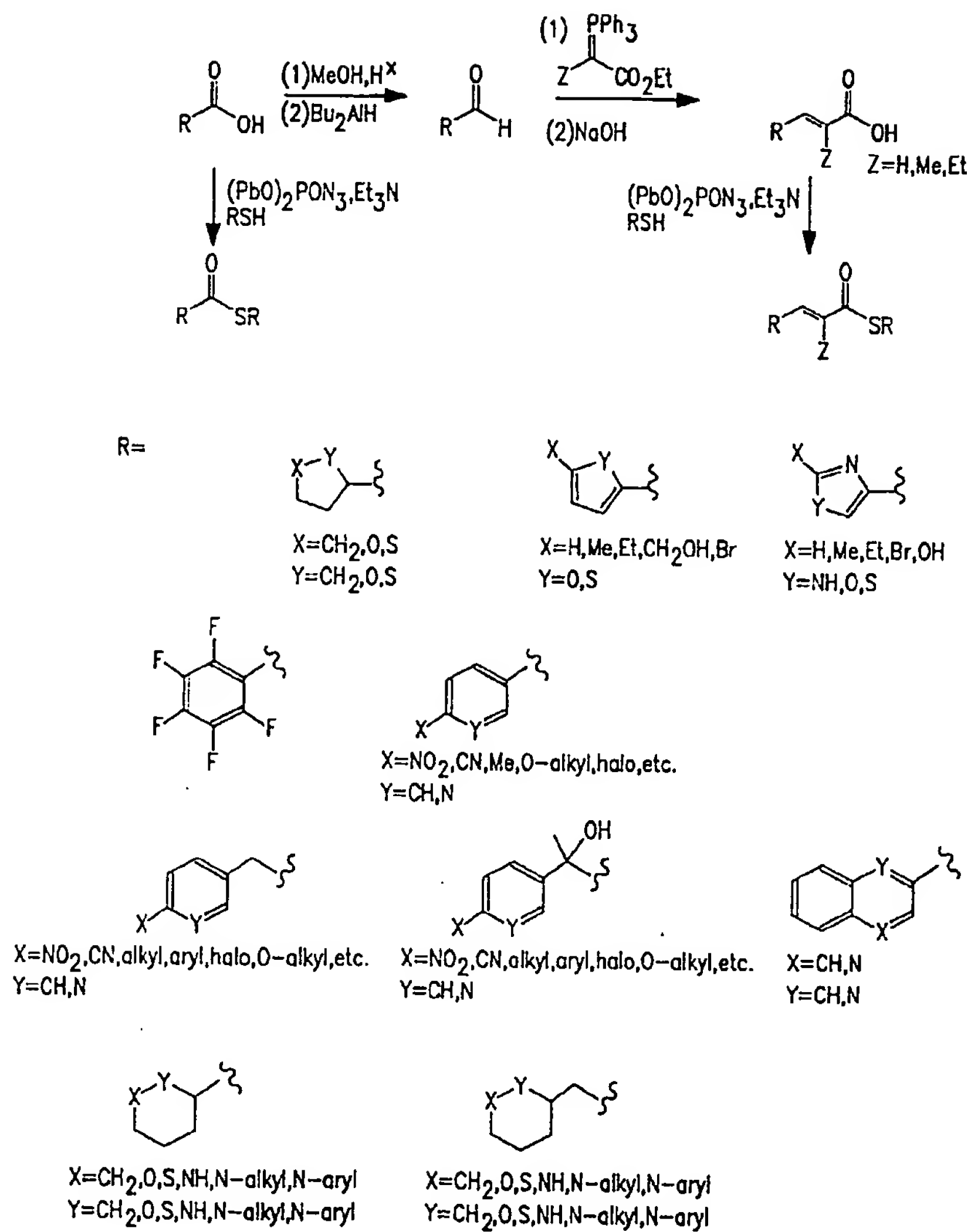


FIG. 2

Alternative Primers for Biosynthetic Epothilone Analogs

SUBSTITUTE SHEET (RULE 26)

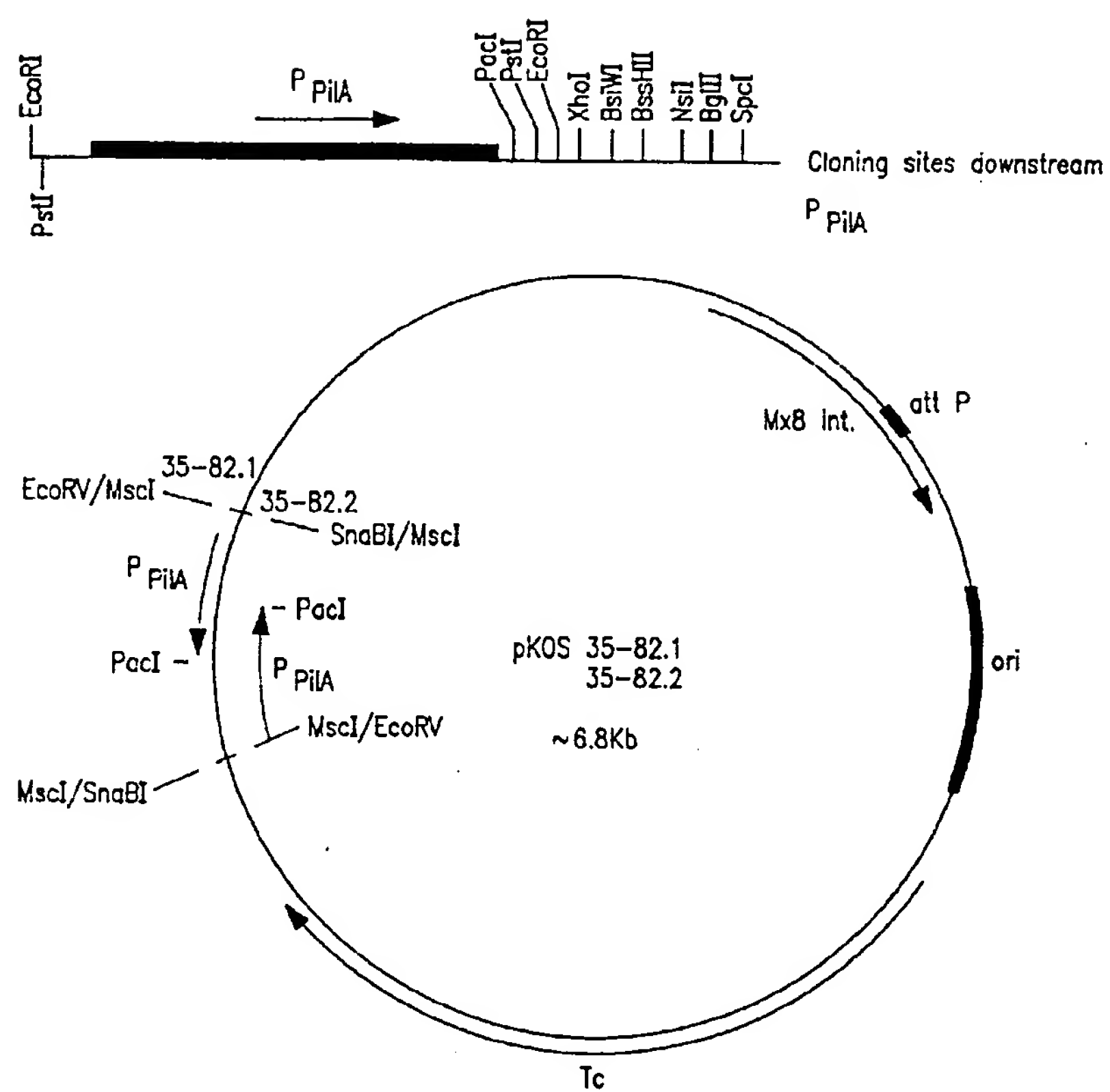


FIG. 3

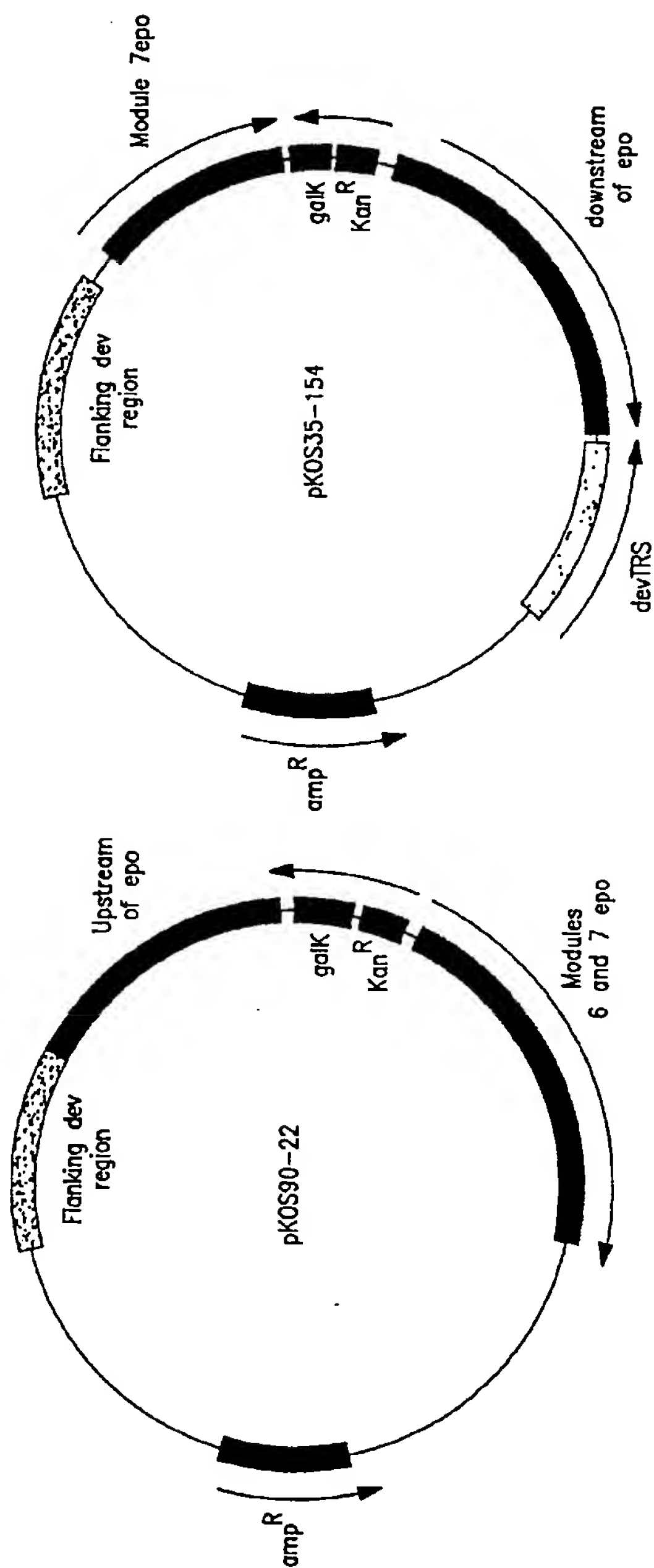


FIG. 4

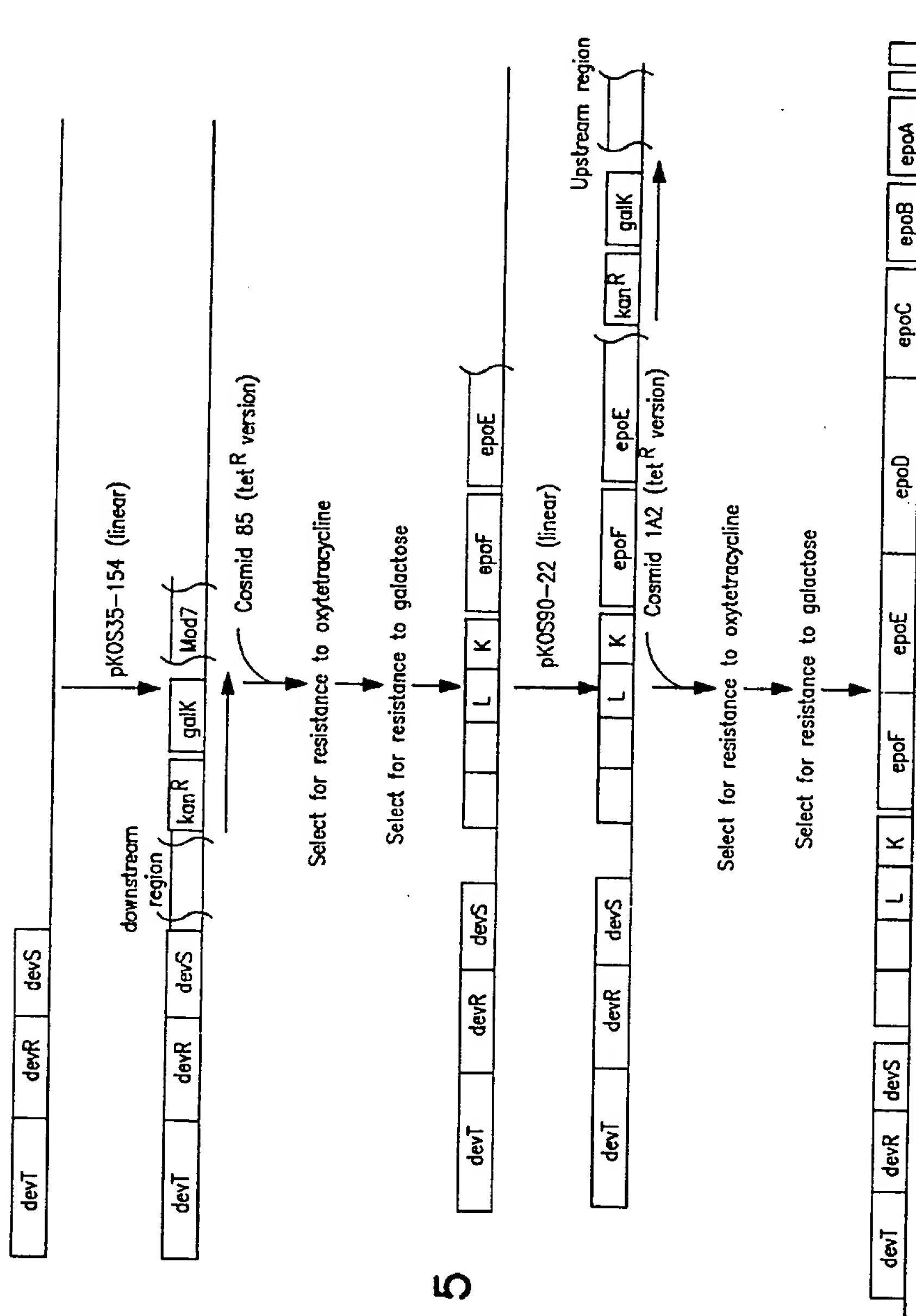


FIG. 5

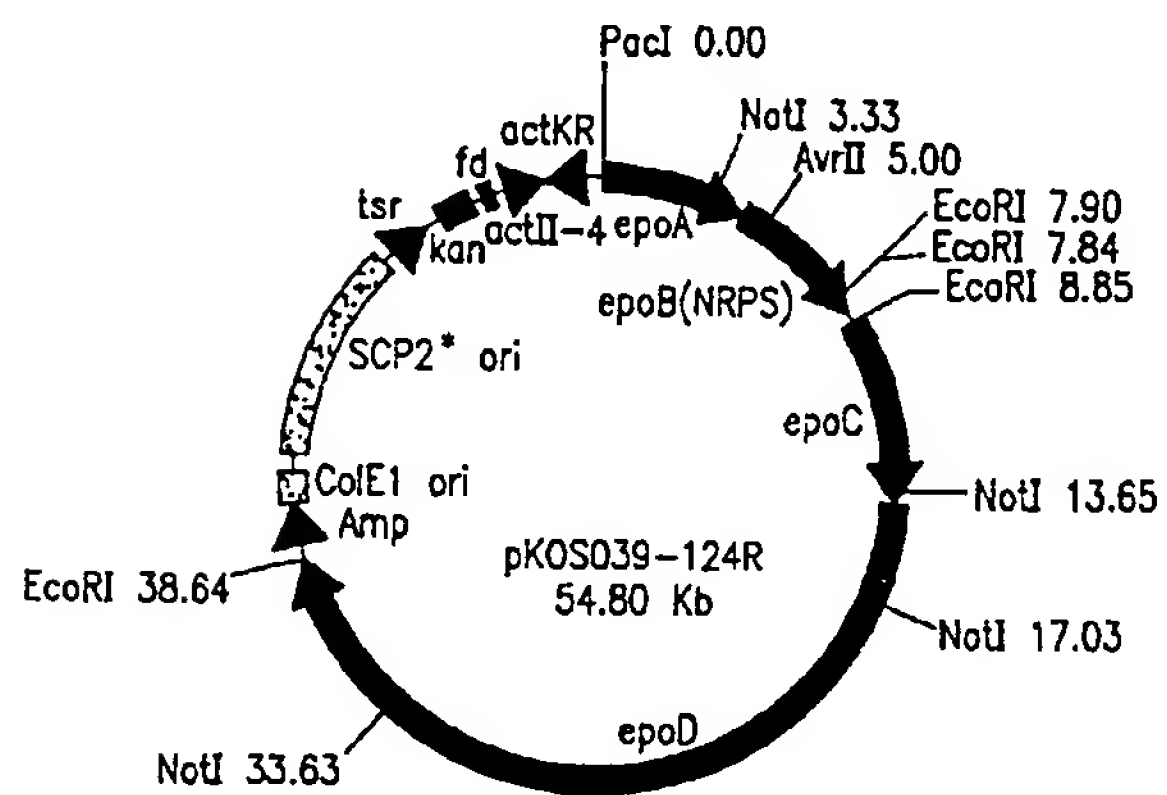
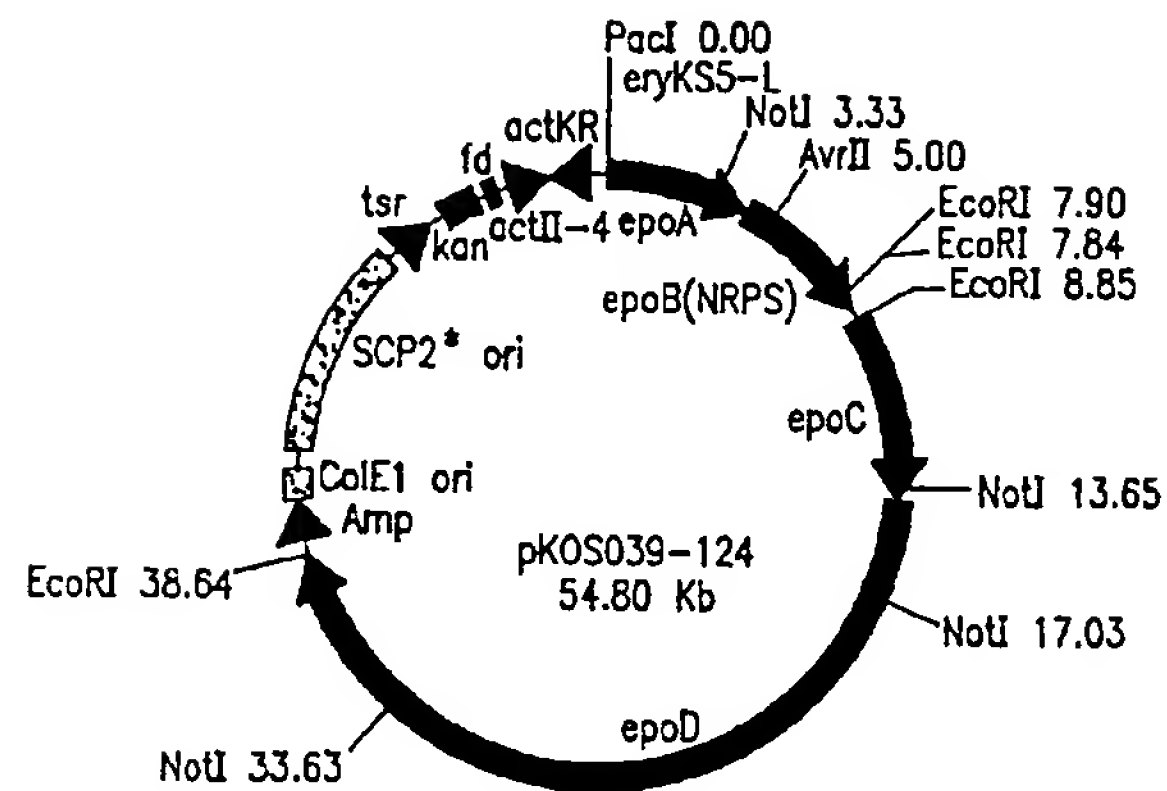


FIG. 6

7/8

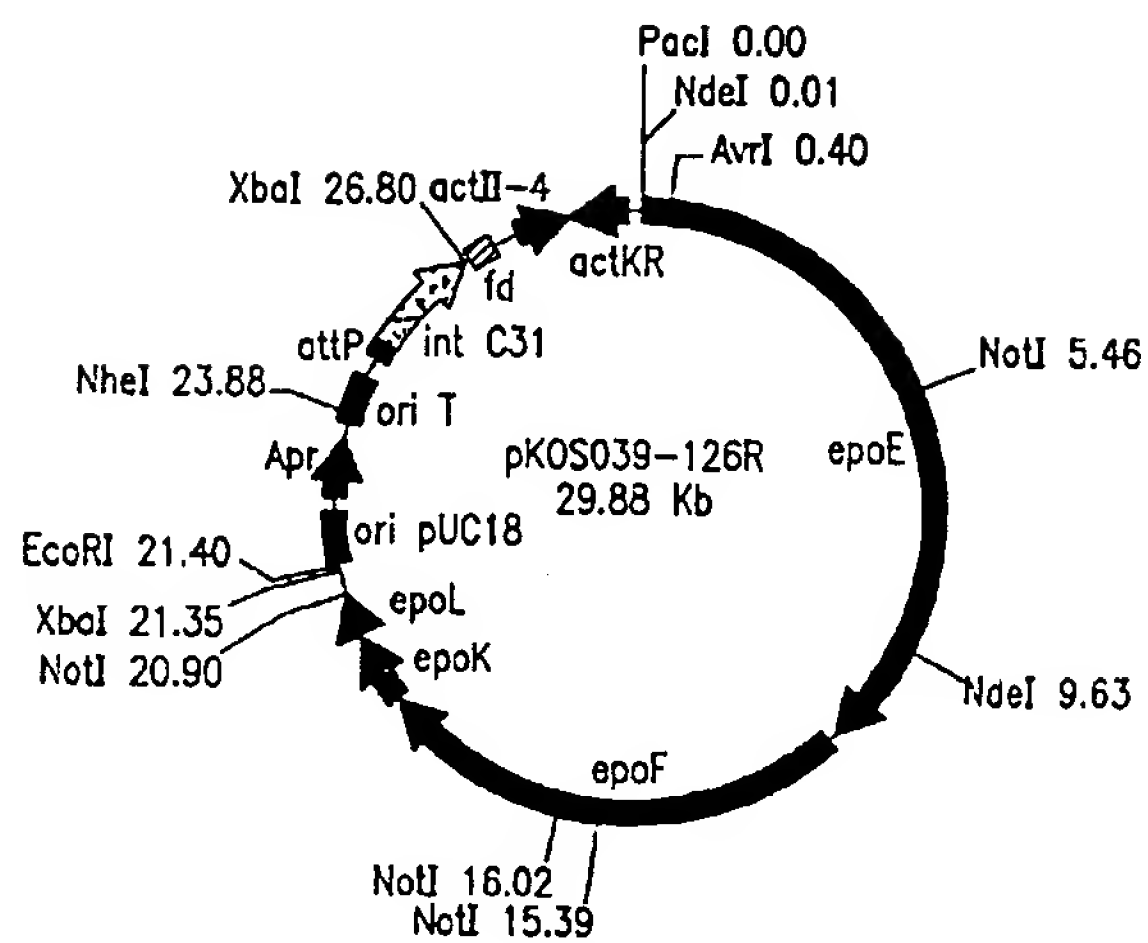


FIG. 7

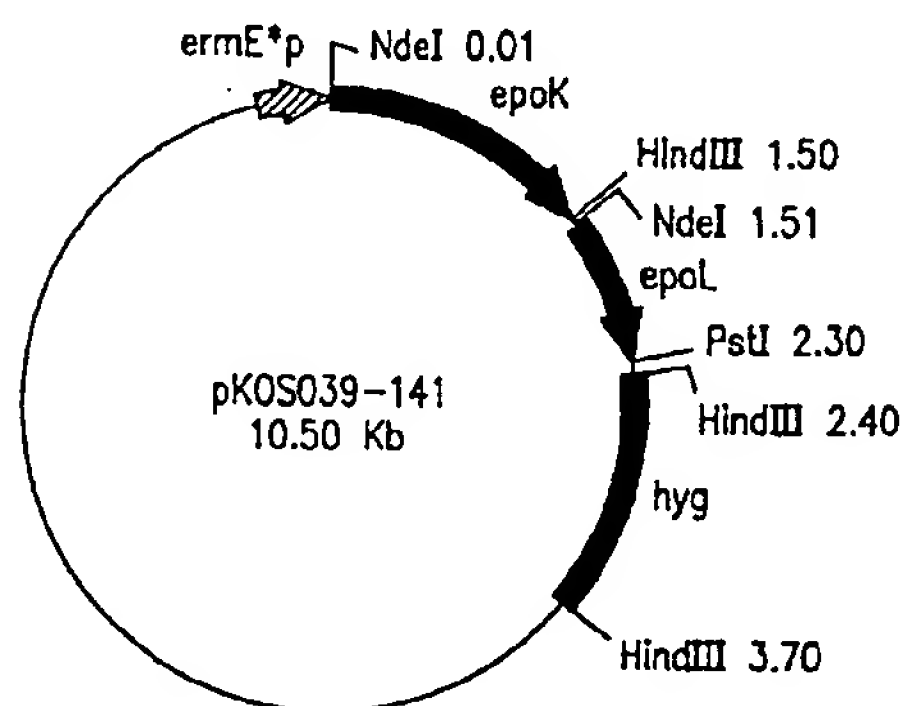


FIG. 8

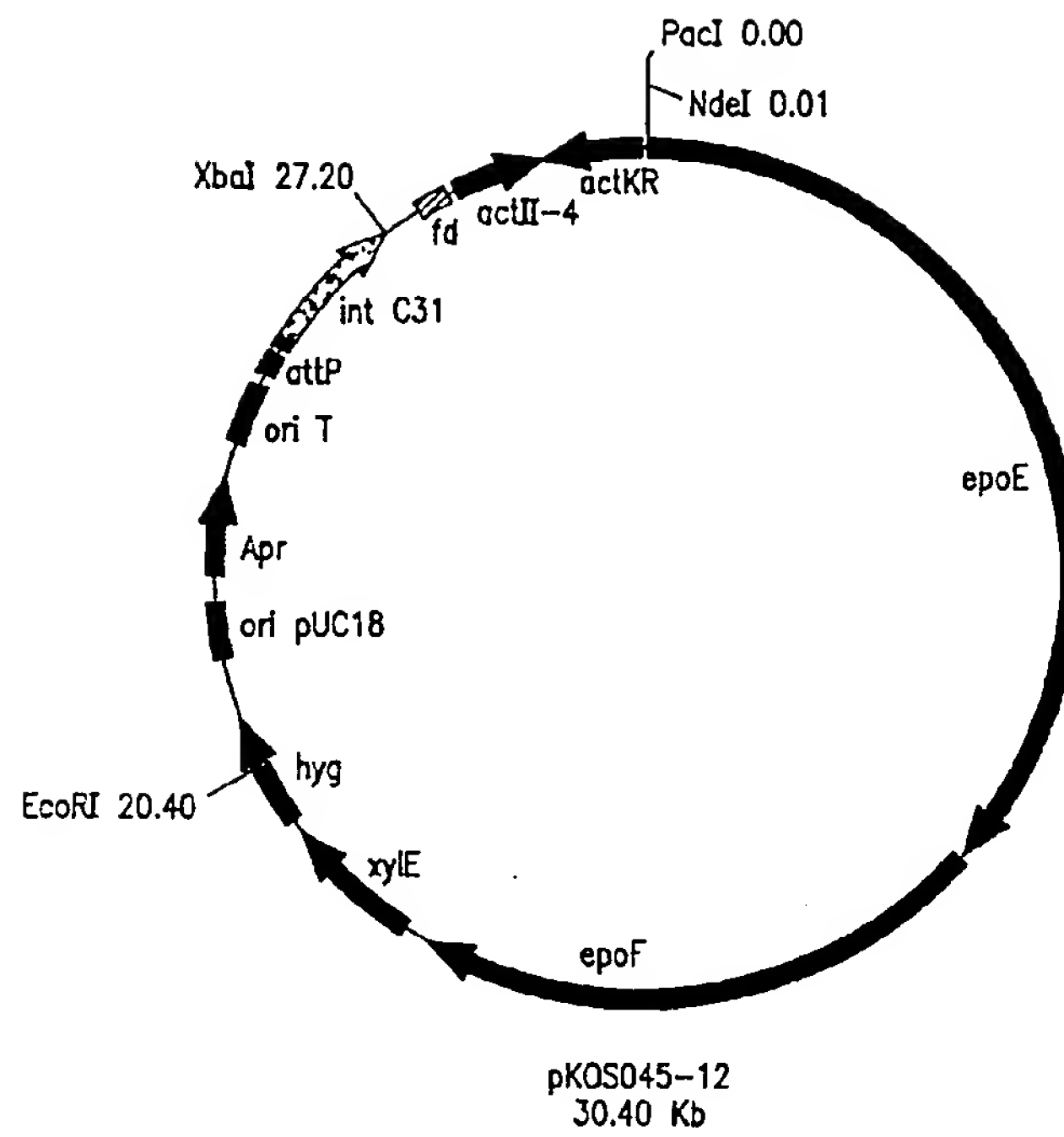


FIG. 9

5
13981 TTCGCCGGCG GTGGCGAGGC GACGCCGTGG GTGCCCGTGG AAGTGGGCTC GCTGCGGCTC
14041 TTGCAGCGGC CTTGGGGGGA GCTGTGGTGC CATGCGCGCG TCGTGAACCA CGGGCGCCAA
14101 ACCUCCGATC GGCAGGGCGC CCACTTTTGG GTGGTCGACA GCTCGGGTGC AGTGGTCGCC
5 14161 GAAGTCAGCG GSCCTCGTG3C GCAGCGGCTT CCGGGAGGGG TCGCGCGGCG CGAAGAAGAC
14221 GATTGGTTCC TSGAGCTC3A GTGGGAACCC GCAGCGGTCG GCACAGCCAA GGTCAACGCG
14281 GGCCGGTGGC TCCTCCTCGG CGGGCGGCGT GGGCTCGGCG CCGCTTTCG CTGATGCTG
10 14341 GAGGCCGGCG GCCATGCCGT CGTCCATGCG GCAGAGAGCA ACACGAGCGC TGCCGCCGTA
14401 CGCGCGCTCC TGGCAAAGGC CTTTGACGGC CAGGCTCCGA CGGCGGTGGT GCACCTCGGC
14451 AGCCTCGATG GGGGT3GCGA GCTCGACCCA GGGCTCGGGG CGCAAGGCGC ATTGGACGCG
10 14521 CCCCAGAGCG CCGACGTCAG TCCCAGTGCC CTCGATCCGC CGCTGGTACG TGGCTGTGAC
14581 AGCGTGCTCT GCACCGTGCA GGCCCTGGCC GGCATGGGCT TTCGAGACGC CCGCGATTG
14641 TGECTTCTGA CCGCGGCGC ACAGGCCGTG GCGCGCGGCG ACGTCTCCGT GACACAGGCA
15 14701 CCGTGCTGG GGCTGGGCGG CGTCATCGCC ATGGAGCAGC CGGATCTGCG CTGCGCTCGG
14761 GTCGACCTCG ATCCGACCGG GCCCGATGGG GAGCTCGGTG CCCTGCTGGC CGAGCTGCTG
15 14821 GCCGACGACG CUGAAGCGGA AGTCGCGTTC CGCGGTGGCG AGCGATGCGT CGCTCGGATC
14881 GTCCCCCGGC AGCCCGAGAC CCGGCCCCGG GGGAGGATCG AGAGCTGCGT TCCGACCGAC
14941 GTCACCATCC GCGCGGACAG CACCTACCTT GTGACCGGCG GTCTGGGTGC GCTCGGTCTG
15001 AGCGTGGCCG GATGGCTGGC CGAGC3CGGC GCTGGTCACC TGCTGCTGGT GGGCCGCTCC
20 15061 GGCGCGGCGA GCGTGGAGCA ACGGGCAGCC GTCGCGGCGC TCGAGGCCCG CGGCGCGCGC
15121 GTCACCGTGG CGAAGGACGA TGTCGCGGAT CCGCGCGCAG TCGAGCGGAT CCTCCGCGAG
20 15181 GTTACCACGT CCGGGATGCC GCTGCGGGGC GTCTGTCATG CCGCCGGCAT CTGGGACGAC
15241 GGGCTGCTGA TGCAGCAGAC TCCCGCGCGG TTTCGTAAGG TGATGGCGCC CAAGGTCCAG
15301 GGGGCTTTC ACCTGCACGC GTTACACGCG GAAGCGCGCG TTTCCTTCTT CGTGCTGTAC
15361 GCTTCGGGAG TAGGGCTCTT GG3CTCGCCG GGCCAGGGCA ACTACGCGCG GGCCAACACG
25 15421 TTCTTCGACG CTCTGGCGCA CCACCGGAGG GCGCAGGGCG TGCCAGCGTT GAGCGTCGAC
15481 TGGGGCTCTT TCCCGAGGT GGGCATGCGG GCCGCGCAGG AAGATCGCGG CGCGCGGCTG
25 15541 GTCTCCCGCG GAATGCGGAG CCTACCCCC GACGAGGCGG TGTCGCTCTT GGCACGCTG
15601 CTTCGAAAGCG GCCGCGTGCA GGTGGGGTGA ATGCCGCTGA ACCCGCGGCT GTGGGTGGAG
15661 CTCTACCCCG CGGCGGCGTC TTGCGGAATG TTGTCGCGCC TGGTGACGGC GCATCGCGCG
30 15721 AGCGCGGGCG GGCCAGCGCG GGACGGGGAC CTGCTCCGCC GCCTCCCTGC TGCCGAGCCG
15781 AGCGCGCGGA GCGGGCTCCT GGAGCCGCTC CTCCGCGCGC AGATCTCGCA GGTGCTGCGC
15841 CTCCCCGAGG GCAAGATCGA GGTGGACGCC CCGCTACAGA GCCTGGGCAT GAACTCGCTG
30 15901 ATGGGGCTCG AGCTCGGCAA CCGCATCGAG GGCATGCTGG GCATCACCGT ACCCGCAACG
15961 CTGTTGTGGA CCTATCCAC GGTGGCGGGC CTGAGCGGGC ATCTGGCGCG GGAGGCATGC
35 16021 GAAGCCGCTC CTGTGGASTC ACCGACACCC ACCGCGGATT CTGCTGTGCA GATCGAGGAG
16081 ATGTGCGAGG ACGATCT3AC GCAGTTGATC GCAGCAAAAT TCAAGGCGCT TACATGACTA
16141 CTCGCGGTCC TACGGCACAG CAGAATCCGC TGAAACAAGC GGCCATCATC ATTCAGCGCC
16201 TGGAGGAGCG GCTCGCTGGG CTCGCACAGG CCGAGCTEGA ACGGACCGAG CCGATCGCCA
35 16261 TCCTCGGTAT CGGCTGCCGC TTCCCTGGCG GTGCGGACGC TCCGGAAGCG TTTTGGGAGC
16321 TGCTCGACGC GAGCGCGGAC GCGGTCCAGC CGCTCGACAG GCGCTGGGCG CTGGTAGGTG
40 16381 TCGCTCCCGT CGAGGCCGTG CCGCACTGGG CCGGGCTGCT CACCGAGCCG ATAGATTGCT
16441 TCGATGCTGC GTTCTTCGGC ATCTCGCCTC GGGAGGCGCG ATCGCTCGAC CCGCAGCATC
16501 GTCTGTTGCT GGAGGTGCTT TGGGAGGGCG TCGACGACGC CGGTATCCCG CCGCGGTCCA
16561 TCGACGGGAG CCGCACCGGT GTGTTCGTG GCGCTTTAC GCGGACTAC GCGGACCGG
45 16621 TCGCTCGGTT GCCGCGCGAG GAGCGAGACG CGTACAGCGC CACCGGCAAC ATGCTCAGCA
16681 TCGCCGCGCG ACGGCTGTG TACACGCTGG GGCTGCGAGG ACCTGCGCTG ACCGTCGACA
16741 CGGCGTGCTC GTCATCGCTG GTGGCGATTG ACCTGCGCTG CCGCAGCCTG CCGCGAGGAG
16801 AGAGCGATCT CCGGTTGGCG GGAGGGGTCA GCACGCTCCT CTCCCCCGAC ATGATGGAAG
16861 CCGCGCGCGC CACGCAAGCG CTGTGCGCCG ATGGTCGTTG CCGGACCTTC GATGCTTCGG
50 16921 CCAACGGGTT CGTCCGTGGC GAGG3CTGTG SCCTGGTCTG CCTCAAACGG CTCTCCGACG
16981 CGCAACGGGA TGGCGACCGC ATCT3GGCGC TGATCCGGGG CTCGCCCCATC AACCATGATG
17041 GCCGGTCGAC CCGGTTGACC GCGCCCAACG TGCT3GCTCA GGAGACGGTC TTGCGCGAGG
45 17101 CGCTGCGGAG CCGCCACGTC GAAGCTGGCG CCGT3GATTA CGTCCGAGAC CACGSAACAG
17161 GGACCTCGCT GGGCGATCCC ATCGAGGTG AGGCGCTGCG GGCGACGGTG GGGCCGGCGC
55 17221 GCTCCGACGG CACACGCTGC GTGCTGGGCG CCGTGAAGAC CAACATCGGC CATCTCGAGG
17281 CCGCG3CAGG CGTAGCGGGC CTGATCAAGG CAGCGCTTTC GCTGACGCAC GAGCGCATCC
17341 CGAGAAACCT CAACTTCCGC ACGCTCAATC CCGGATCCG GCTCGAGGGC ACCGCGCTCG
17401 CGTTGGCGAC CGAGCCGGTG CCGTGGCCGC GCACGGACCG TCCGCGCTTC GCGGGGTGA
50 17461 GCTCGTTCGG GATGAGCGGA ACGAACGCGC ATGTGGTGCT GGAAGAGGCG CCGCGGTGG

5 17521 AGCTGTGGCC TGCCGCGCCG GAGCGCTCGG CGGAGCTTTT GGTGCTGTCT GGCAAGAGCG
17581 AGGGGGCGCT CGACGCGCAG GCGGCGCGGC TGC3CGAGCA CCTGGACATG CACCCGGAGC
17641 TCGGGCTCGG SGACGTGUCG TTCAGCCTGG CGACGACCGC CAGCGCGATG ACCCACCAGC
17701 TCGCCGTCCC SGTGACGTCG CGCGAGGGGC TGCTGGCGGC GCTTTCGGCC GTGGCGCAGG
5 17761 GGCAGACGCC GCGGGGGGCG GCGCGCTGCA TCGCGAGCTC CTCGCGCGGC AAGCTGGCGT
17821 TGCTGTTCAC CGGACAGGGC GCGCAGACGC CGGGCATGGG CCGGGGGCTC TCGCGGCGT
10 17881 GGCCAGCGTT CCGGAGGGCG TTCGACCGGT GCGTGACGCT GTTCGACCGG GAGCTGGACC
17941 GCGCGCTGCG CGAGGTGATG TGGGCGGAGG CCGGGAGCGC CGAGTCGTTG TTGCTGGACC
18001 AGACGGCGTT CACCCAGCCC GCGCTCTTCG CGGTGGAGTA CCGCTGACG GCGCTGTGGC
10 18061 GGTCTGTGGG CGTAGAGCCG GAGCTCCTGG TTGGGCATAG CATCGGGAG CTCGTGGCGG
18121 CGTGCGTGGC GGGGCTCTCG GCGGGCGGCG CGATGGTGTG GCTCGTGGCG GCGCGCGGCG
18181 GGCTGATGCA GGGGCTCTCG GCGGGCGGCG CGATGGTGTG GCTCGTGGCG CCGGAGGCGG
15 18241 AEGTGGCCCG GCGGTGGCG CCGCAGCGCG CGTGGCTCTC GATCGCGGCG GTCAATGGGC
18301 CCGAGCAGGT GGTGATCGCG AAGCGGTGCA AAGCGGTGCA GCGATCGCG SCGGGGTTCG
15 18361 CCGCGCGCGG CGTGCGCACC AAGCGGTGCA AAGCGGTGCA GCGATCGCG SCGGGGTTCG
18421 TCGAACCAGT GCTGGAGGAC TTCGGGCGGG TGGCGGCGTC GGTGACGTAC CCGCGGCCAA
18481 GCGTTCGCT GGTGAGCAAC CTGAGCGGGA AGGTGGTCAC GGACGAGCTG AGCGCGCCGG
18541 GCTACTGGGT GCGGCACGTG CCGGAGCGCG TGGCGTTCGC GACGCGGGTG AAGCGCGTGC
18601 ACGAAGCCCG CCGGGGACAG TTCCTCGAAG TGGGCCCCGA GCCGACGCTG CTCGGCCTGT
20 18661 TGCCASCCTG CCGGCGGAG GCGGAGCGCA CGTTGCTGGC GTCGTTGCGC GCGGGCGCG
18721 AGGAGGCTGC GGGGGTGTCT GAGGCGCTGG GCGAGCTGTG GGCCGCTGGC GGCTCGGTCA
18781 GCTGGCCGCG CGTCTTCCCC ACGGCTGGGC GCGGGGTGCG GCTGCGGACC TATCCGTGGC
18841 AGCGGCAGCG GTACTGGATC GAGGCGCCCG CCGAACGGCT CCGAGCCACG GCCGCGGATG
20 18901 CGCTGCCCGA GTGGTCTCTAC CCGGTGGACT GGCCCGAGAT GCCTCGCTCA TCCGTGGATT
25 18961 CCGGCGGAGC CCGGTCCGCG GGTGCTGCG TGGTGGCGCA CCGGGGTGGA GTCGGGAGG
19021 CCGCCGCGGC GCGGCTTTCT TCGCAGCGAT GTTCGTGCGC CGTGTCCAT GCGCCCGCG
25 19081 AGGCTCCGC GGTGCGCGAG CAGGTGACCC AGGCTCTCG TGGCCGCAAC GACTGGCAGG
19141 GGTGCTGTA CCGTGGGGT CTGGACCGCG TCGTGAGGC GGGGCGATCG GCCGAAGAGG
19201 TCGGCAAGT CACCCATCTT GCCACGCGCG CCGTGTCTCG GCTGATTAG GCGGTGGGCA
30 19261 CCGGGCCGCG CTCACCCCGG CTCTGGATCG TGACCCGAGG GGCTGACAG GTGGGCGCG
19321 AGCCTGACGC TGCCCCCTGT CAGCGCGCGC TGTGGGTAT GGGCCGGTC GCGGCGCTGG
19381 AGCATCCCG CTCTGGGGC GCGCTCGTG ACCTGGATCC GGAGGAGAG CCGACGGAGG
30 19441 TCGAGGCCCT GGTGGCCGAG CTGCTTTCTG CCGACGCGCA CGATCAGCTG GCATTCCGCC
19501 AGGGGCGCG GCGCGCAGCG CCGCTCGTG CCGCCCCACC GGAGGGAAAC GCAGCGCCCG
35 19561 TGTCGCTGTC TGCGGAGGG AGTTACTTGG TGACGGGTGG GCTGGGCGCC CTTGGCCTCC
19621 TCGTTGCGCG GTGGTTGGTG GAGCGCGGG CCGGGCACCT TGTGCTGATC AGCCGGCAGC
19681 GATTGCCCCA CCGCGAGGAA TGGGGCCGAG ATCAGCGCGC AGAGGTGCGC GCGCGCATTC
19741 CCGCGATCGA GCGGCTGGAG GCGCAGGGCG CCGGGGTAC CGTGGCGCG GTGACGTGG
35 19801 CCGATGCCGA AGGCATGCCG GCGCTCTTGG CCGCCGTCGA GCGCCGCTG CCGGGGGTGG
19861 TGCACGCCG GGTCTGTCT GACGACGGGC TGCTGGCCCA CCAGGACGCC GGTGCGCTCC
19921 CCGGGGTGTT GCGCCCCAAG GTGGAGGGG CATGGGTGCT GCACACCCTT ACCCGCGAGC
19981 AGCCGCTGGA CCTCTTCGTA CTGTTTTCTT CCGGCTCGGG CGTCTTCGG TCGATCGGCC
20041 AGGGCAGCTA CCGGCGAGGC AATGCCITTT TGGACGCGCT GCGGACCTC CGTCGAACGC
45 20101 AGGGGCTCGC CGCCCTGAGC ATCCCTGGG GCCTGTGGG GGAGGGGGGG ATGGGCTCGC
20161 AGGCGCAGCG CCGGGAACAT GAGGCATCG GAATCTGGG GATGCCGAG AGTCGTGCC
40 20221 TGGCGGCGAT GGAATGGCTG CTCGGTACGC CCGGACGCA GCGCGTGGT ATCCAGATGG
20281 ATTGGGCCCC TGCGGGAGCG GCTCCGCGCG ACGCGAGCG AGGCGGCTT TGGGATCGGC
20341 TGGTAACTGT CACGAAAGCG GCCTCCTCCT CGGCCGTGCC AGCTGTAGAG CGCTGGCGCA
50 20401 ACGCGTCTCT TGTGGAGACC CGCTCGGCG TCTACGAGCT TGTGCGCGG GTGGTCCCG
20461 GGTGATGGG CTTTACCGAC CAAGGCACGC TCGACGTGCG ACGAGGCTTC GCGGACAGG
20521 GCCTCGACTC CCGATGCTCT GTGGAGATCC GCAAACGGCT TCAGGGTGAG CTGGGTATGC
45 20581 CGCTGTCSGC GACGCTGCGG TTCGACCATC CGACCGTGG GCGGCTGGT GAATACTTGC
20641 TGAGCCAGCG GCTGGAGCTG CAGGACCGCA CCGACGTCC AAGCGTTCGG TTGCCGCGA
20701 CAGAGGACCC GATCGCCATC GTGGGTGCG CCTCCGCTT CCGGGCGGG GTGAGGACC
55 20761 TGGAGTCTTA CTGGCAGCTG TTGACCGAG GCGTGGTGG CAGCACCAG GTGCCGGCG
20821 ACCGCTGGAA TGGGCGAGC CCGCGCGGCC CCGGCTCGGG AGAGGCTCC AGACAGACCT
20881 ACGTGCCCG GGGTGGCTTT CTGCGGAGG TGGAGACGTT CGATGCGGCG TTCTTCCACA
20941 TCTCGCCTCG GGAGGCGATG AGCCTGGACC CGCAACAGCG GCTCTGCTG GAAGTGAGCT
50 21001 GGGAGGCGAT CGAGCGCGCG GGCCAGGACC CGTCGGCGCT GCGCGAGAGC CCCACGGCG

5 21061 TGTTCGTGGG CGCGGGCCCC AACGAATATG CCGAGCGGGT GCAGGACCTC GCCGATGAGG
21121 CCGCGGGGCT CTACAGCCGC ACCGGCAACA TGCTCAGCCT TGCGGCGGGA CGGCTGTCAI
21181 TTTTCCTGGG CCTGCACGGG CCGACCCTGG CTGTGGATAC GGCGTGCTCC TCCTCGCTCG
21241 TGGCGCTGCA CCTCGGCTGC CAGAGCTTGC GACGGGGCGA GTGCGACCAA GCCCTGGTTG
5 21301 GCGGGGTCAA CATGCTGCTC TCGCCGAAGA CCTTCGCGCT GCTCTCACGG ATGCACGCGC
21361 TTTCGCCCCG CGGGCGGTGC AAGACGTTCT CGGCCGACGC GGACGGCTAC GCGCGGGCCG
10 21421 AGGGCTGCGC CGTGGTGGTG CTCAAGCGGC TCTCCGACGC GCAGCGCGAC CGCGACCCCA
21481 TCCTG3CGGT GATCCGGGGT ACGGCGATCA ATCATGATGG CCGAGCAGC GGGCTGACAG
21541 TGCCAGCGG CCTGCCCCAG GAGGCGCTGT TACGCCAGGC GCTGGCGCAC GCAGGGGTGG
10 21601 TTCCGGCCGA CGTCGATTC GTGGAATGCC ACGGGACCG GACGGCGCTG GCGGACCCGA
21661 TC3AGGTGCG GCGCTGAGC GACGTGTACG GGCAAGCCCG CCCTGCGGAC CGACCGCTGA
21721 TCCTGGGAGC CGCCAAGGCC AACCTTGGGC ACATGGAGCC CGCGGCGGGC CTGGCCGGCT
15 21781 TGCTCAAGGC GGTGCTCGCG CTGGGGCAAG AGCAAATACC AGCCCAGCCG CAGCTGGGCG
21841 AGCTCAACCC GCTCTTCCCG TGGGAGCGCG TGCCGGTGGC GGTGGCCCGC GCAGCGGTGC
15 21901 CGTGGCCCGC CACGGACCGT CCGCGCTTCG CGGGGGTGAG CTCTTCGGG ATGAGCGGAA
21961 CGAACGCGCA TGTGGTGCTG GAAGAGGCGC CGCGGTCCA GCTGTGGCCT GCCCGCCCG
22021 AGCGCTCGGC GGAGCTTTTG GTGCTGTGCG GCAAGAGCGA GGGGGCGCTC GACGCGCAG
22081 CGGCGCGGCT GCGCAGCAC CTGGACATGC ACCCGAGCT CGGGCTCGGG GACGTGCGCT
22141 TCAGCCTGGC GACGACGCGC AGCGCCATGA ACCACCGCT CGCGGTGGCG GTGACGTGCG
20 22201 GCGAGGGGCT GCTGGCGGCG CTTTCGGCCG TGCGCGAGGG GCAGACGCGC CCGGGGGCGC
22261 CGCGCTGCAT CGCGAGCTCG TCGCGCGGCA AGCTGGCGTT CCTGTTCACC GGACAGGGCG
22321 CGCAGACGCC GGGCATCGGC CGGGGGCTTT GCGCGGCGTG GCCAGCCTC CGAGAG3CGT
22381 TCGACCGGTG CGTGGCGCTG TTCGACCGGG AGCTGGACCG CCCGCTGTGC GAGGTGATGT
22441 GGGCGGAGCC GGGGAGCGCC GAGTCGTTGT TGCTCGACCA GACGGCGTTC ACCCAGCCCG
25 22501 CGCTCTTCAC GGTGGAGTAC GCGCTGACGG CGCTGTGGCG GTCTGGGGC GTAGAGCCCG
22561 AGCTGGTGGC TGGGCATAGC GCCGGGGAGC TGGTGGCGGC GTCCGTGGCG GGGGTGTTCT
25 22621 CGCTGGAAGA TGGGTGAGG CTCGTGGCGG CGCGCGGGCG GCTGATGCAG GGGCTCTCGG
22681 CGGGCGGCGC GATGGTGTG CTCGGAGCGC CGGAGGCGGA GGTGCGCGC GCGGTGGCGC
22741 CGCACGCGGC GTGGGTGTG ATCGCGCGG TCAATGGGCC GGASCAGGTG GTGATCGCGG
30 22801 GCGTGGAGCA AGCGGTGCAG GCGATCGCG CGGGGTTCGC GCGCGCGGC GTGCGACCA
22861 AGCGGTGCA TGTCTCGCAC GCATCCCACT CGCGCTGAT GGAACCGATG CTGGAGGAGT
22921 TCGGCGGGGT GCGGCGCTCG CTGACGTACC GCGGCGCAAG CGTTTCGCTG GTGAGCAACC
30 22981 TGAGCGGGAA GGTGCTCACG GACGAGCTGA GCGCGCCGGG CTACTGGGTG CCGCACGTGC
23041 GGGAGCGGCT GCGCTTCGCG GACGGGGTGA AGGCGCTGCA CGAAGCCGGC GCGGGGACGT
35 23101 TCCTCGAAGT GGGCCCGAAG CCGACGCTCG TCGGCTGTG GCCAGCTTGC CTGCGGAGG
23161 CGGAGCCGAC GCTGCTGGCG CCGGGCGCGA CGGGCGCGA GGAGGCTCCG GGGGTGCTCG
23221 AGGCGCTGGG CAGGCTGTGG GCCCGCGCG GCTCGGTGAG CTGGCGGGC GTCTTCCCCA
23281 CGGCTGGGCG GCGGGTGCCG CTGCGACCT ATCGGTGGCA GCGGCGAGCG TACTGGCCCG
35 23341 ACATCGAGCC TGACAGCCGT CGCCACGCG CCGCGGATCC GACCCAAGGC TGGTCTATC
23401 GCGTGGACTG GCGGAGATA CCTCGAGCC TCCAGAAATC AGAGGAGGCG AGCCGCGGGA
23461 GCTGGCTGGT ATTGGCGGAT AAGGGTGGAG TCGGCGAGGC GGTGCTGCA GCGCTGTGCA
23521 CACGTGGAAT TCCATGCGTC GTGCTCCATG CGCCGCGAGA GACATCCGCG ACCGCGAGC
23581 TGGTGACCGA GGCTGCCGCG GGTGGAAGCG ATTGCGAGGT AGTGCTCTAC CTGTGGGTC
45 23641 TGGAC3CCGT CGTCGCGCG GAGGCGTGA TCGATGACAT CGGCGACGCG ACCGTCGTG
23701 CTACC3CGCC GGTGCTCGGC TTGGCTCGGT TTCTGAGCAC CGTGTCTTGT TCGCCCGAC
40 23761 TCTGGSTCGT GACCCGGGGG GCATGCATG TTGSCGACGA GCCTGCGATC GCCCCTTGT
23821 AGGCGGCGTT ATGGGGCATG GCGCGGTGAG CGCGGCTGCA GCA7CCCGGG GCCTGGGGCG
23881 GGCTCGTGA CCTGGATCCC CGAGCGAGCC CGCCCAAGC CAGCCCGATC GACGGCGAGA
50 23941 TGCTCGTAC CGAGCTATTG TCGCAGGAGA CCGAGGACCA GCTCGCCTTC CGCATGGGC
24001 GCGGCGACGC GGCACGCTG GTGGCCGCC CGCCACGGGG GGAAGCGGCA CCGCGCTCGC
24061 TGCTGCGGA GCGAGCTAC CTGGTGACGG GAGGCTCGG TGGGCTGGGC CTGATCGTGG
45 24121 CCCAGTGGCT GGTGGAGCTG GGAGC3CGGC ACTTGGTGT GACCAGCCGG CGCGGGTTC
24181 CCGACCGGCA GCGGTGGCG GAGCAGCAGC CGCTGAGAT CCGCGCGCG ATCGCAGCG
24241 TCGAGGCGCT GGAGGCGCG GGTGCACGG TGACCGTGGC AGCGGTGGAC GTGGCCGACG
55 24301 TCGAACCAGT GACAGCGCTG GTTTCGTGCG TCGAGCCCC GCTGCGAGGG GTGGTGCACG
24361 CCGCTGGCGT CAGCGTCATG CGTCCACTGG CGGAGACGGA CGAGACCTG CTCGAGTCG
24421 TGCTCCGTCC CAAGGTGGCC GGGAGCTGGC TGCTGCACCG GCTGCTGCAC GCGCGGCTC
24481 TCGACCTGTT CGTCTGTTT TCGTGGGCG CAGCGGTGT GGGTAGCCAT AGCCAGGGTG
50 24541 CGTACGCGGC GGCCAACGCT TTCTCGACC GGCTCGCGCA TCTTCGGCGT TCGCAATCGC

5 24601 TGCCTGCCTT GAGCGTCGCG TGGGGTCTGT GGGCCGAGGG AGGCATGGCG GACCCGAGAG
24661 CTCATGCACG TCTGAGCGAC ATCGGGGTTT TCCCCAATGTC GACGTCGGCA GCGTTGTCCG
24721 CGCTCCAGCG CTTGGTGGAG ACCGGCGCGG CTCAGCGCAC GGTGACCCGG ATGGACTGGG
24781 CGCGCTTCGC GCCGGTGTAC ACCGCTCGAG GCGCTCGCAA CCTGCTTTTC GCGCTGGTTCG
5 24841 CAGGGCGCGA CATCATCGCG CCTTCCCTC CGGCGGCAGC AACC CGGAAC TGGCGTGCC
24901 TGTCCGTTGC GGAAGCCCGC ATGGCTCTGC ACGAGGTCGT CCATGGGGCC GTCGCTCGGG
10 24961 TGCTGGGCTT CCTCGACCG AGCGCGCTCG ATCCTGGGAT GGGGTTCAAT GAGCAGGCC
25021 TCGACTCGTT GATGGCGGTG GAGATCCGCA ACCTCCTTCA GGCTGAGCTG GACGTGCGGC
25081 TTTTCGACGAC GCTGGCCTTT GATCATCCGA CGGTACAGCG GCTGGTGGAG CATCTGCTCG
10 25141 TCGATGTACT GAAGCTGGAG GATCGCAGCG ACACCCAGCA TGTTCGGTCG TTGGCGTCAG
25201 ACGAGCCCAT CGCCATCGTG CGAGCCGCTT GCGGCTTCCC GGGCGGGGTG GAGGACCTGG
25261 AGTCCCTACTG GCAGCTGTTG GCCGAGGCGG TGGTGGTCAG CGCCGAGGTG CCGCCGACC
15 25321 GGTGGGATGC GCGGACTGG TACGACCTG ATCCGGAGAT CCCAGGCCGG ACTTACGTGA
25381 CCAAAGGCGC CTTCTGCGC GATTTGCAGA GATTGGATGC GACCTTCTTC CGCATCTCGC
15 25441 CTCGCGAGAG GATGAGCCTC GACCCGAGC AGCGGTTGCT CTGGAGSTA AGCTGGGAGG
25501 CGCTCGAGAG CCGCGGTATC GCTCCGATA CGCTCGAGA TAGCCCCACC GGGGTGTTTCG
25561 TGGGTGCGCG GCCCAATGAG TACTACACGC AGCGGCTGCG AGGCTTCACC GACCGAGCGG
25621 CAGGGCTGTA CCGCGGCACC GGAACATGC TCAGCGTTGC GGCTGGACGG CTGTCTTTT
20 25681 TCCTGGGTCT GCACGGCCCG ACGTGGCCA TGGATACGGC GTGCTGCTCC TCCCTGGTCC
20 25741 CGCTGCACCT CGCCTGCCAG AGCTTGCAG TGGGCGAGTC CGATCAAGCG CTGGTTGGCG
25801 GGTCAACGT GCTGCTCGCG CCGGAGACCT TCGTCTGCT CTCACGGATG CGCGCGCTTT
25861 CCGCCSACGG GCGGTCAAG ACGTCTCGG CCGAGGCGGA CGGCTACGG CCGGGCGAGG
25921 GGTGC3CCGT GGTGGTGTCT AAGCGGCT3C CCGATGCGCA GCGC3CCGG GACTCCATCC
25981 TGGCCCTGAT CCGGGGAAGC GCGGTGAACC ACGACGGCCC GAGCAGCGGG CTGACCGTGC
25 26041 CCAACGGACC CGCCAGCAA GCATTGTCTG GCCAGGCGCT TCGCAAGCA GCGGTGTCTC
26101 CGGTGACGCT TGATTTTGTG GAGTGTACG GACAGGGAC GCGCTGGGC GACCCGATCG
25 26161 AGGTGCAGGC GCTGAGCGAG GTGTATGGTC CAGGGCGCTC CGAGGATCGA CCGCTGCTGC
26221 TGGGGGCCGT CAAGGCCAAC GTCCGCGATC TGGAGGCGGC ATCCGGCTTG GCCAGCCTGC
26281 TCAAGGCCGT GCTTGCGCTG CCGCACGAGC AGATCCCGSC CCAGCCGAG CTGGGGGAGC
30 26341 TCAACCCGCA CTTGCCGTGG AACAC3CTGC CCGTGGCGST GCCACGTAAC GCGGTGCCST
26401 GGGGGCGCGG CGCACGGCCG CGTCC3GCCG GCGTGAGCGC GTTCGGGTTC AGCGGAACCA
26461 ACGTGATGT CGTGCTGGAG GAGGCACCG AGGTGGAGCT GGTGCCCGCG GCGCCGGCGC
30 26521 GACCGGTGGA GCTGGTGTG CTATCGGCCA AGAGCCCGGC GCGCTGGAC GCGCGGGCGC
26581 AACGGCTCTC GGCGCACCTG TCCGCGCACC CCGAGCTGAG CCTCGGCGAC GTGGCGTTCA
35 26641 GCCTGGCGAC GACGCGCAGC CCGATGGAGC ACCGGCTCGC CATCGCGACG ACCTCGCGCG
26701 AGGCCCTGCG AGGCGCGCTG GACGCCCGG CGCAGCGGCA GACGCCCGAG GGCGCGGTGC
26761 GCGGCAAGGC CGTGTCTCA CGCGGTAAGT TGGCTTTCTT GTTCA3CGGA CAGGGCGCGC
26821 AAATGCCGGG CATGGCCGT GGGCTGTAC AGGCGTGGCC AGCGTTCCGG GAGGCGTTCC
35 26881 ACCGGTGCCT CCGCTCTTC GATCGGGAGC TCGACAGCC TCTGCGCGAG GTGATGTGGG
26941 CTGCGCCGGG CCTCGCTCAG GCGCGCGGCG TCGATCAGAC CGCGTACCG CAGCCGGCTC
40 27001 TCTTTGCGCT GGAGTACCG CTGGCTGCCC TGTGGCGTTT GTGGGGCGTG GAGCCGACG
27061 TACTCCTCGG TCATAGCATC GCGGAGCTGG TCGCCGCTG CGTGGCGGGC GTGTTCTCGC
27121 TCGAAGACGC GGTGAGGTTG GTGGCCGCGC GCGGGCGGCT GATGCAGGCG CTGCCCCCGC
27181 GCGGTGCCAT GGTGCGCATC GCAGCGTCCG AGGCGGAGGT GGCGGCTCC GTGGACCCC
45 27241 ACGCCGCCAC GGTGTGATC GCGCGGCTCA ACGGTCTGA CGCGTCTGT ATCGCTGGCG
27301 CCGAGGTACA GGTGCTCGCC CTCGCGCGCA CGTTCGCGG GCSTGGGATA CGCACGAAGA
40 27361 GGCTCGCCGT CTCCTATCG TTCCACTCGC CGCTCATGGA TCCGATGCTG GAAGACTTCC
27421 AGCGGGTCGC TCGGAGATC GCGTACCGCG CGCCAGACCG CCCGGTGGTG TCGAATGTCA
27481 CCGGCCACGT CGCAGGCCCC GAGATCGCCA CGCCGAGTA TTGGGTCCGG CATGTGCGAA
50 27541 GCGCCGTGCG CTTGGCGAT GGGGCAAGG CGTTGCATGC CCGGGTGGC GCCACGTTCC
27601 TCGAGATTGG CCCGAAGCG GTCTGTCTG GGCTATTGCC AGCGTGCTC GGGGAAGCGG
27661 ACGCGGTCTT CGTGCCGTG CTACGCGCG ACCGCTCGGA ATGCGAGGTG GTCTCGCGG
45 27721 CGCTCGGGAC TTGGTATGCC TGGGGGGTG CCTCGACTG GAAGGGCGTG TTCCCGATG
27781 GCGCGCGCCG CGTGGCTCTG CCCATGTATC CATGGCAGCG TGAGCGCCAT TGGATGGACC
55 27841 TCACCCCGCG AAGCGCCGCG CCTGCAGGGA TCGCAGGTG CTGGCCGCTG GCTGGTGTG
27901 GGCTCTGCAT GCGCGCGCT GTGTTGCCA ACGTGTCTC GATCGGACCA CGCCATCAGC
27961 CCTTCTCGG TGATCACCTC GTGTTGGCA AGGTGGTGGT GCGCGGGCG TTTTCATGTCG
28021 CGGTGATCCT CAGCATCGCC GCCGAGCGCT GCGCCGAGCG GCGGATCGAG CTGACAGGCG
50 28081 TGGAGTTCTT GAAGGCGATC GCGATGGAGC CCGACCA3GA GGTGAGCTC CAGCCGCTGC

5
28141 TCACCCCGA AGCCGCCGGG GATGGCTACC TGTTCGAGCT GGCGACCCTG GCGGCGCCGG
28201 AGACCGAACG CCGATGGACG ACCCAGCCCC GCGGTCGGGT GCAGCCGACA GACGGCGCGC
28261 CCGGCGCGTT GCCGCGCCTC GAGGTGCTGG AGGACCGCGC GATCCAGCCC CTCGACTTCG
28321 CCGGATTCTT CACAGGTTA TCGGCGGTGC GGATCGGCTG GGGTCCGCTT TGGCGATGGC
5 28381 TCGAGGACGG GCGCGTCGGC GACGAGGCCT CGCTTGCCAC CCTCGTGCCG ACCTATCCGA
28441 ACGCCACGA CGTGGCGCC TCGACCCGA TCCTGCTGGA CAACGGCTTT GCGGTGAGCC
10 28501 TGCTGGCAAC CCGGAGCGAG CCGGAGGACG ACGGACGCC CCCGCTGCCG TCGCCGTTG
28561 AACGGGTGCG GTGGTGGCG GCGCGGTTG GAAGGGTGC GTGTGGCGEC GTGCGCGGT
28621 CGCAGGATT CGGTGTCTCG AGCTTCGTGC TGGTCGACGA AACTGGCGAG GTGGTCGCTG
10 28681 AGGTGGAGGG ATTTGTTTGC CGCCGGGCGC CGCGACAGGT CTTCCTGCGG CAGGAGTCGG
28741 GCGCGTCCAC TGCAGCCTTG TACCGCCTCG ACTGGCCCGA AGCCCCCTTG CCCGATGCGC
28801 CTGCGGAACG GATGGAGGAG AGCTGGGTGC TGGTGGCAGC ACCTGGCTCG GAGATGGCCG
15 28861 CCGCGCTCGC AACACGCTC AACCGCTCG TACTCGCCGA ACCCAAAGGC CTCGAGGCGG
28921 CCCTCGCGGG GGTGTCTCCC GCAGGTGTA TGTGCTCTG GGAACCTGGA GCCCAGAGG
15 28981 AAGCTCCGGC GCGGCGCGAG CGTGTGGCGA CCGAGGGCCT TTCGGTGGTG CAGGCGCTCA
29041 CCGATCGCGC GGTGCGCCTG TGGTGGGTGA CCACGGGCGC CGTGGCTGTC GAGGCGGTC
29101 AGCGGGTGCA GGTGCGCACA GCGCCGGTAT GGGGCTGGG CCGGACAGTG ATGACGAGC
29161 GCGCGAGCT CAGCTGCACT CTGGTGGATT TGGAGCCGGA GGTGATGCC GCGCGTTCAG
20 29221 CTGACCTCT GCTGCGGGAG CTCGGTCGCG CTGACGACGA GACCCAGGTG GTTTCCGTT
29281 CCGGAGAGCG CCGCTAGCG CGGCTGTCGA AACGACAAC CCGCAAGGG CTCTTGGTCC
29341 CTGACGAGA ATCCTATCGA CTGGAAGGTG GCGAGAAGGG CACATTGGAC CAGCTCCGCC
29401 TCGCGCGGC ACAGCGCCGG GCACCGGCG CCGGCGAGGT CGAGATCAAG GTAACCGCT
29461 CCGGGCTCAA CTCCGGACC CTCCTCGCTG TGCTGGGAAT GTATCCGGGC GACGCTGGGC
25 29521 CGATGGGCGG AGATTGTGCC GGTATCGTCA CCGCGGTGGG CCAGGGGGTG CACCACCTCT
29581 CGGTGCGCGA TGTGTCTATG ACGTGGGGA CGTTGCATCG ATTCGTCACG GTCGACGCGC
25 29641 GGCTGCTCGT CCGGAGCCT GCAGGGCTGA CTCGCGCGA GGCAGCTACG GTGCCGGTTG
29701 CGTTCTGAC GGCTGGCTC GCTCTGACG ACCTGGGGAA TCTGCGGCGC GGCGAGCGGG
29761 TGCTGATCCA TGCTGCGGCC GCGGCGGTG GCATGGCCGC GGTGCAAATC GCGCATGGA
29821 TAGGGGCGCA GGTGTTCCGC ACAGCGAGCC CGTCCAAGTG GGCAGCGGTT CAGGCCATGG
30 29881 GCGTCCGCG CACGACATC GCCAGCTCGC GGACGCTGGA GTTTGCTGAG ACGTTCCGGC
29941 AGGTACCCGG CCGCCGGGGC GTGGACGTGG TGCTCAACGC GCTGGCCGGC GAGTTCGTGG
30001 ACGCGAGCCT GTCCCTGCTG ACGACGGCG GCGGTTTCTT CGAGATGGGC AAGACCGACA
30 30061 TACGGGATCG AGCCGCGGTG CCGGCGGCC ATCCCGGTGT TCGTATCGG GTATTGACA
30121 TCCTGGAGCT CGCTCCGGAT CGAATCGAG AGATCCTCGA GCGCGTGGT GAGGGCTTTG
35 30181 CTGCGGACA TCTGCGCGCA TTGCGGTGTC ATGCGTTCGC GATCACCAG GCGGAGCGAG
30241 CGTTTCGGTT CATGCGCGAA GCGCGCATC AGGGCAAGGT CGTGCTCTG CCGGCGCCCT
30301 CCGCAGCGCC CTTGGCGCGC ACGGCAACG TACTGCTGAC CGGTGGGCTG GGAGCGTTGG
30361 GGCTCCACGT GGCCCGCTGG CTCGCCCAGC AGGGCGCGCC GCACATGGTG CTCACAGGTC
35 30421 GCGGGGCGCT GGATACGCCG GCGCGTGCCA AAGCCGTGCG GGAGATCGAA GCGCTCGGCG
40 30481 CTCGGGTGAC GATCGCGCGC TCGGATGTC CCGATCGGAA CCGCTGGAG GCTGTGCTCC
30541 AGGCCATTCC GGCGGAGTGG CCGTTACAGG GCGTGATCCA TGCAGCCGGA GCGCTCGATG
30601 ATG3TGCTG TGATGAGCAG ACCACCGACC GCTTCTCGCG GGTGCTGGCA CCGAAGGTGA
30661 CTG3CGCCTG GAATCTGCAT GAGCTCACGG CCGCAACGA TCTCGCTTTC TTCGTGCTGT
45 30721 TCTCCTCCAT GTCGGGGCTC TTGGGCTCGG CCGGGCAGTC CAACTATGCG GCGGCCAACA
30781 CCTTCCTCGA CCGCTGSCC GCGCATCGGC GGGCCGAAGG CCTEGCGGCG CAGAGCCTCG
40 30841 CGTGGGGCCC ATGGTCCGAC GGAGGCATGG CAGCGGGGCT CAGCGCGGCG CTGCGGCGC
30901 GGCTCGCTCG GCATGGGATG GGAGCGCTGT CGCCGCTCA GGGCACCGCG CTGCTCGGCC
30961 AGGCGCTGGC TCGGCCGAA ACGCAGCTCG GGGCGATGTC GCTCGACGTG CGTCCGGCAA
50 31021 GCGAAGCTTC GGGAGCGGCA GTGCCGCTG TGTGGCGCGC GCTGGTGGC GCGGAGGCGC
31081 GCGATGCGGC GGCTGGGGCG CAGGGGGCAT TGGCCGCGCG CCTTGGCGCG CTGCCGAGG
31141 GCGTTCGCGC GACGAGGTG CGCAAGGTG TGCAGGCCGA GATCGCGCGC GTGCTTTCAT
45 31201 GGGGCGCGCG GAGCCCGTG CCCGTCGATC GCGCGCTGTC GGAATTGGGC CTCGACTCGC
31261 TCACGGCGGT GGAGCTGCGC AACCTCTCG GCCAGCGGT GGGTGGGACG CTGCCGCGA
31321 CGCTGGCATT CGATCACCCG ACGGTCGACG CGCTACGCG CTGGCTGCTC GATAAGGTCC
55 31381 TGGCCGTGGC CGAGCCGAGC GTATCGCCCG CAAAGTCGTC GCGCAGGTC GCCCTCGACG
31441 AGCCCATTCG GGTGATCGGC ATCGGCTGCC GTTTCGAGG CCGCGTGACC GATCCGAGT
31501 CGTTTTGGCG GCTGCTCGAA GAGGGCAGCG ATGCCGCTGT CAGGCTGCGC CATGAGCGAT
31561 GGGACATCGA CCGTTCTAT GATCCGGATC CGGATGTGCG CCGCAAGATG ACGACACGCT
50 31621 TTGGCGGCTT CCTGTCCGAT ATCAGCCGCT TCGAGCCGCG CTTCCTCGGC ATCTCGCCCG

5 31681 GCGAAGCGAC GACCATGGAT CCGCAGCAGC GGCTGCTCCT GGAGACGAGC TGGGAGGCGT
31741 TCGAGCGCGC CGGGATTTTG CCCGAGCGGC TGATGGGCAG CGATACCGGC GTGTTCGTGG
31801 GGCTCTTCTA CCAGGAGTAC GCTGCGCTCC CCGGCGGCAT CGAGGCGTTC GATGGCTATC
5 31861 TAGGCACCGG CACCACGSCC AGCGTCGCCT CGGGCAGGAT CTCTTATGTG CTCGGGCTAA
31921 AGGGGCCGAG CCTGACGGTG GACACCGCGT GCTCCTCGTC GCTGGTCGCG GTGCACCTGG
31981 CCTGCCAGGC GCTGCGGCGG GCGGAGTGT? CGGTGGCGCT GGCCGCGCGC GTGGCGCTGA
10 32041 TGCTCACGCC GCGACGTTT GTGGAGTTCA GCGGCTGCG AGGCCTGGCT CCCGACGGAC
32101 GGTCGAAGAG CTTCTCGGCC GCAGCCGACG GCGTGGGGTG SAGCGAAGGC TGGCCCATGC
10 32161 TCCTGCTCAA ACCGCTTCCG GATGCTCAGC GCGATGGGGA TCCGATCCTG GCGGTGATCC
32221 GCGGCACCGC GGTGAACCA GATGGGCGCA GCAACGGGCT GACGGCGCCC AACGGTCTGT
32281 CGCAGCAAGA GGTGATCCGT CCGGCCCTGG AGCAGGCGGG GCTGGCTCCG GCGGACGTCA
32341 GCTACGTCGA GTGCCACGGC ACCGGCACGA CGTTGGGCGA CCCCATCGAA GTGCAGGCCC
15 32401 TSGGCGCCGT GCTGGCACAG GGGCGACCTT CCGACCGGCC GCTCGTGATC GGGTCGGTGA
32461 AGTCCAATAT CGGACATACG CAGGCTGCGG CCGGCGTGCG CCGTGTATC AAGGTGGCGC
15 32521 TGGCGCTCGA GCGCGGGCTT ATCCGAGGA GCCTGCATTT CGACGCGCCC AATCCGCACA
32581 TTCCGTGCTC GAGCTCGCC GTGAGGTGG CCGCAAACC CGTCGATGG ACGAGAAACG
32641 GCGCGCCGCG ACGAGCCGGG GTGAGCTCGT TTGGCGTCAG CCGGACCAAC GCGCACGTGG
20 32701 TGCTGGAGGA GCGGCCAGCG GCGGCGTTCC GCGCCGCGGC GGCGCCTTCA GCGGAGCTTT
32761 TCGTGCTGTC GCGAAGAGC GCGCGGGCGC TGGACGCGCA GGCGGCGCGG CTTTCGCGCC
20 32821 ATGTGCTTGC GCACCCGGAG CTCGGCTCG GCGACCTGGC GTTCAGCCTG GCGACGACCC
32881 GCAGCCCGAT GACGTACCGG CTCGCGGTGG CCGCGACCTC GCGCGAGGCG CTGTCTGCGG
32941 CGCTCGACAC AGCGGCGCAG GGGCAGGCGC CCGCCGAGC GGCTCGCGGC CACGCTTCCA
33001 CAGGCAGCGC CCAAAGGTG GTTTTCCTCT TTCTTGGCCA GGGCTCCAG TGGCTGGGCA
25 33061 TGGGCCAAAA GCTCCTCTCG GAGGAGCCCG TCTTCCGCGA CCGGCTCTCG GCGTGTGACC
33121 GAGCGATTCA GGCCGAAGCC GGCTGGTCCG TGCTCGCCGA GCTCGCGGCC GATGAGACCA
25 33181 CCTCGCAGCT CCGCCCGCATC GACGTGGTGC AGCCGCGCCT GTTCGCGATC GAGGTGCGCG
33241 TGTCGGCGCT GTGGCGGTCT TGGGGCGTGC AGCCGATGC AGTGGTAGGC CACAGCATGG
33301 GCGAAGTGGC GGCCGCGCAC GTCGCGCGCG CCGTGTGCT CGAGGATGCT GTAGCGATCA
33361 TCTGCCGGCG CAGCCTGCTG CTGCGCGGGA TCAGCGGCCA AGGCGAGATG GCGGTCTGCG
30 33421 AGCTCTCCCT GGCCGAGGCC GAGGACGCGC TCCTGGGCTA CCAAGATCGG CTCAGCGTGG
33481 CGGTGAGCAA CAGCCC3CGA TCGACGGTGC TGGCGGGCGA GCGGCGAGCG CTCGCGAGAG
33541 TGCTGGCGAT CCTTGCGGCA AAGGGGTGT TCTGCGTGC AGTCAAGGTG GACGTGCGCA
30 33601 GCCACAGCCC ACAGATCGAC CCGTGCAGC ACGAGCTATT GGCAGCATTG GCGGAGCTCG
33661 AGCCGCGACA AGCCAGCTG TCGATGCGCT CGACGGTGAC GAGCAGATC GTGGCGGGCC
35 33721 CGGAGCTCGT GGCGAGCTAC TGGGCGGACA ACGTTCGACA GCGGTGCGC TTCGCGGAAG
33781 CGGTGCAATC GTTGATGGA GCGGTGATG GGCTGTTCGT GGAGATGAGC CCGCATCCGA
33841 TCCTGACGAC GTCGGTTCGAG GAGATCCGAG GGGCGACGAA GCGGGAGGGA GTGCGGCTGG
33901 GCTCGTTGCG GCGTGGACAG GACGAGCGCC TGTCATGTT GGAGGCGCTG GGAGCGCTCT
35 33961 GGGTACACGG CCAGGCGGTG GCGTGGGAGC GGCTGTCTC CCGGGCGGC GCGGGCTTCC
34021 GTGCGGTGCC GCTGCCGACC TATCCCTGGC AGCGCGAGCG GTACTGCGTC GAAGCGCCGA
34081 CCGGCGGGCG GCGGAGCGGC AGCGGCTTTG CTCATGCGGG CAGTACCCG CTCCTGGGTG
34141 AAATGCGAGC CCTGTGACCC CAGAGGAGCA CCGCGGTGTG GGAGACGACG CTGGATCTCA
34201 AACGGCTGCC GTGGCTCGGC GATCACCGGG TGCACGGGGC GGTGCTGTTT CCGGGCGCGG
34261 CGTACCTGCA GATGGCGCTT TCGTCTGGGG CCGAGGCTT GGGTGACGGT CCGCTCCAGG
45 34321 TCAGCGATGT GGTGCTCGCC GAGGCGCTGG CTTTCGCGA TGATACGCCG GTGGCGGTGC
34381 AGGTGATGGC GACCGAGGAG CCAACAGGCC GCCTGCAATT CCACGTTGCG AGCCGGGTGC
40 34441 CCGGCCACGG CCGTCTGACC TTTCGAAGCC ATGCCCGCGG GGTGCTGCGC CAGACCGAGC
34501 GCGCCGAGGT CUCGGCGAGG CTGGATCTGG CCGGCTTTCG TGCCCGGCTT CAGGCCAGCG
34561 CACCCGCTGC GGCTACCTAT GCGGCGCTGG CCGAGATGGG GCTCGAGTAC GGCCAGCGCT
50 34621 TCCAGGGGCT TGTCGAGCTG TGGCGGGGGG AGGGCGAGGC GCTGGCACGT GTGCGGCTCC
34681 CCGAGGCCGC CCGCTCCCA GCGCGGTGCC GGCTCCACCC CCGCTCTTTG SATGCGTGCT
45 34741 TCCACGTGAG CAGC3CCTT GCTGACCGCG SCGAGGCGAC GCCATGGGTA CCGTTCGAAA
34801 TCGGCTCGCT GCGGTGCTT CAGCGGCCCT CCGGGGAGCT GTGGTGTCT GCGCGGAGCG
34861 TGAGCCACGG AAAGCCAACA CCGATCGGC GGAGTACCGA CTTTTGGGTG GTGACAGCA
55 34921 CCGGCGCGAT CCGCGCCGAG ATCTCCGGGC TCGTGGCGCA GCGGCTCGCG GGAGGTGTAC
34981 GCGGCGCGCA AGAAGACGAC TGGTTCATGG AGCCGGCTTG GGAACCGACC GCGGTCCCG
35041 GATCCGAGGT CACGGCGGGC CCGTGGCTGC TCATCGGCTC GGGCGGCGGG CTCGGCGCTG
35101 CGCTCTACTC GCGCTGACG GAAGCTGGCC ATTCCGTCT CCAC3CGACA GGGCACGGCA
50 35161 CGAGCGCCGC CCGGTTGCAG GCACTCCTGA CCGGCTCCTT CGACGGCCAG GCGCCGACGT

5 35221 CCGTGGTGCA CCTCGGCAGC CTCGATGAGC GTGGCGTGCT CGACGCGGAT GCGCCCTTCG
35281 ACGCCGATGC CCTCGAGGAG TCGCTGGTGC GCGGCTGCGA CAGCGTGCTC TGGACCGTGC
35341 AGCCCGTGGC CGGGGCGGGC TTCCGAGATC CTCGCGGTT GTGGCTCGTG ACACGCGGGC
5 35401 CTCAGGCCAT CGGCGCGGGC GACGTCTCCG TGGCGCAACC GCCGCTCCTG GGGCTGGGGC
35461 GCGTTATCGC CTTGGAGCAC GCCGAGTSC GCTGCGCTCG GATCGACCTC GATCCAGCGC
35521 GCGCGGACGG ASAGGTCGAT GAGCTGCTTG CCGAGCTGTT GGCCGACGAC GCCGAGGAGG
10 35581 AAGTCGCGTT TCGCGGCGGT GAGCGGCGCG TGGCCCGGCT CGTCCGAAGG CTGCCGAGA
35641 CCGACTGCCG ASAGAAAATC GAGCCCGCGG AAGCCCGGCC GTTCCGGCTG GAGATCGATG
35701 GGTCCGGCGT GCTCGACGAC CTGGTGCTCC GAGCCACGGA GCGGCGCCCT CCTGGCCCGG
10 35761 GCGAGGTCGA GATCGCCGTC GAGGCGCGGG GGCTCAACTT TCTCGACGTG ATGAGGGCCA
35821 TGGGGATCTA CCCTGGGCCC GGGGACGGTC CCGTTGCGCT GGGCGCCGAG TGCTCCCGCC
35881 GAATTGTCGC GATGGCCGAA GGTGTCGAGA GCCTTCGTAT CCGCCAGGAC GTCGTGGCCG
35941 TCGCGCCCTT CAGTTTCGGC ACCACGTC AATCGACGC CCGATGGTC GCACCTCGCC
15 36001 CCGCGGCGCT GACGGCCGCG CAGGCGCCG CCGTGGCCGT CGCATTCATG ACGGCTGGT
15 36061 ACGTCTCGT CCATCTGGGG AGGCTCCGGG CCGGCGAGCG CGTGCTCATC CACTCGCGCA
36121 CCGGGGGGCA CCGGCTCGCT GCTGTGAGA TCGCCCGCCA CCTCGCCGCG GAGATATTTG
36181 CGACCGCTG3 TACGCCGGAG AAGCGGGCGT GGCTGCGCGA GCAGGGGATC GCGCACGTGA
36241 TGGACTCGCG GTCGCTGGAC TTCGCGGAGC AAGTGTGCGC CGCGACGAAG GCGGAGGGGG
20 36301 TCGACGTGCT GTTGAACCTG CTGTCTGGCC CCGGATCGA CCGAGCCTT GCGACCTCG
20 36361 TCGCGGACG3 CCGCTTCATC GAGCTCGGCA AGACG3ACAT CTATGCAGAT CGTCTCTGG
36421 GGTTCGCTCA CTTTAGGAAG AGCCTGTCTT ACAGCGCCGT CGATCTTGGG GGTTCGGCCG
36481 TGGCTCGGCC CGAGCGGCTC GCAGCGCTGC TGGCGGAGGT GGTGGACCTG CTCGACGCGG
36541 GAGCGCTGCA GCCGCTTCG GTAGAGATCT TCCCTCTCTC GCGGGCCGCG GACGCTTCC
25 36601 GGAATAATGGC GCAAGCGCAG CATCTCGGGA AGCTCCTGCT CCGGCTGGAG GACCCGACG
25 36661 TCGCGATCCG CGTTCGCGG GATCCGCGG TCGCCATCCG CCGGACGCG ACCTACCTCG
36721 TGACCGCGG TCTGGGTGGG CTCGGTCTGA GCGTGGCTGG ATGGCTGGCC GACGAGGGG
25 36781 CTGGGCATCT GGTGCTGGT GCGCCCTCCG GTGCGGTGAG CCGGAGCAG CAGACGCTG
36841 TCGCCGCGCT CGAGGCGCAC GCGCGCGTGC TCACGGTAGC GAGGGCAGAC GTCGCGGATC
36901 GGGCGCAGAT CGAGCGGATC CTCCGCGAGG TTACCGCGTC GGGGATGCCG CTCCGCGCGG
30 36961 TCGTTCATGC GCGCGGTATC CTGGACGACG GGCTGCTGAT GCAGCAACC CCGCGCGGCT
37021 TCCGCGCGGT CATGGCGCCC AAGGTCCGAG GGGCCTTGA CCTGCATGCG TTGACACCGG
37081 AAGCGCGGCT CTCCTCTTTC GTGCTGTACG CTTGCGGAGC AGGGCTCTTG GGCTCGCCGG
30 37141 GCCAGGGCAA CTACGCGCGG GCCAACACGT TCCTCGACGC TCTGGCACAC CACCGGAGGG
37201 CGCAGGGGCT GCCAGCATTG AGCATCGACT GGGGCTGTT CGCGGACGTG GGTTCGGCCG
35 37261 CCGGGCAGCA AAATCGCGGC GCACGGGTGG TCACCCGCGG GACGCGGAGC CTCACCCCGG
37321 ACGAAGGGCT GTGGGCGCTC GAGCGTCTGC TCGACGCGCA TCGACCCAG GCGGGGTCA
37381 TGCCGTTTGA CGTGCGGCAG TGGGTGGAGT TCTACCCGGC GCGGGCATCT TCGCGGAGGT
37441 TGTCGCGGCT GGTGACGGCA CCGCGGCTGG CTTCCGCTCG GCTCGCCGGG GATCGGGACC
35 37501 TGCTCGAAGC GCTCGCCACC GCGCGCGTGG GCGCGCGGGC AGGAATGCTG CAGGAGGTG
40 37561 TCGCGCGCGA GGTCTCCAG GTGCTGCGCC TCCCGAAGG CAAGCTCGAC GTGGATGCGC
37621 CGCTACGAG CCTGGGAATG GACTCGCTGA TGGGGCTAGA GCTGCGCAAC CGCATCGAGG
37681 CCGTGTCTCG CATCACCATG CCGGCGACCC TGCTGTGGAC CTACCCACG GTGGCAGCGC
37741 TGAGTGCGCA TCTGGCTTCT CATGTCGTCT CTACGGGGGA TGGGGAATCC GCGCGCCCGC
37801 CCGATACAGG GAACGTGGCT CCAATGACCC ACGAAGTCGC TTCGCTCCAC GAAGACGGGT
45 37861 TGTTCGCTT GATTGATGAG TCACTCGCGC GTGCGGGAAA GAGGTGATTG CGTGACAGAC
37921 CGAGAAGGCC AGCTCCTGGA GCGCTTGGCT GAGGTACTC TGGCCCTTCG CAAGACGCTG
40 37981 AACGAGCGCG ATACCCTGGA GCTCGAGAAG ACCGAGCCGA TCGCATCGT GGGGATCGGC
38041 TGCCGCTTCC CCGGCGGAGC GGGCACTCCG GAGGCGTCTT GGGAGCTGCT CGACGACGGG
50 38101 CCGGACGCGA TCCGGCGGCT CGAGGAGCGC TGGGCGCTCG TAGGTGTGCA CCCAGGCGAC
38161 GACGTACCGC GCTGGGCGGG GCTGCTCACC GAAGCCATCG ACGGCTTCGA CGCGCGTTC
38221 TTCGSTATCG CCCCCGCGG GGCACGGTCC CTCGACCCGC AGCATCGCTT GCTGCTGGAG
38281 GTCGCTGGG AGGGGTTCGA ACACCCCGGC ATCCCGCCTA GGTCCCTCGT CCGGAGCCGC
45 38341 ACCGCGGTGT TCGTCGGCGT CTGCGCCACG GAGTATCTCC ACGCCGCGT CGCGACCCAG
38401 CCGCGCGAAG AGCGGGACGC GTACAGCACC ACCGGCAACA TGCTCAGCAT CGCCGCCGGA
55 38461 CCGCTATCGT ACACGCTGGG GCTGCAGGGA CCTTGCCTGA CCGTCGACAC GCGGTGCTCG
38521 TCATCGCTGG TGGCCATTCA CCTCGCCTGC GCGAGCCTGC GCGCTCGAGA GAGCGATCTC
38581 GCGCTGGCGG GAGGGGTCAA CATGCTTCTC TCCCGGACA CGATGCGAGC TCTGGCGCGC
38641 ACCCAGGCGC TGTCGCCCCA TGGCGGTTGC CAGACCTTCG ACGGCTCGGC CAACGGGTTC
50 38701 GTCCGTGGGG AGGGTGGCG TCTGATCGTG CTCAAGCGAT TGAGCGACGC GCGGCGGGAT

5 38761 GGGGACCGGA TCTGGGCGCT GATCGGAGGA TCGGCCATCA ATCAGGACGG CCGGTGACGG
38821 GGGTTGACGG CGCCCAACGT GCTCGCCAG GGGCGCTCT TCGCGAGGC GCTGCGGAAC
38881 GCGGCGTCC AGGCGAGGC CATCGGTTAC ATCGAGACCC ACCGGGCGGC GACCTCGCTG
5 38941 GCGGACCCCA TCGAGATCGA AGCGCTGCGC ACCGTGGTGG GGCCGGCGCG AGCCGACGGA
39001 GCGGCTGCG TGCTGGGCGC GGTGAAGACC AACCTCGGCC ACCTGAGGG CGCTGCCGGC
39061 GTGGCGGGCC TGATCAAGGC TACACTTTTC CTACATCACG AGCGCATCCC GAGGAACCTC
10 39121 AACTTTTCGTA CCGTCAATCC GCGGATCCGG ATCGAGGGGA CCGGCTCGC GTTGGCGACC
39181 GAACCCGTGC CCTGGCCGCG GACGGGCGCG ACGGCTTCG CCGGAGTGAG CTCGTTCGSG
39241 ATGAGCGGGA CCAACGCGCA TGTGGTGTTC GAGGAGGCGC CCGCGGTGGA GCCTGAGGCC
10 39301 GCGGCCCCCG AGCGCGCTGC GGAGCTGTTT GTCTGTCCG CCAAGAGCGT GCGGCGCTG
39361 GATGCGCAGG CAGCCCGGCT GCGGACCACT CTGGAGAAGC ATGTGAGCT TGGCTTCGCG
39421 GATGTGGCGT TCAGCTGGC GCGGCGCGC AGCGCGATGG AGCACCAGCT GCGGTCGCG
39481 GCGAGCTCGC GCGAGGCGCT GCGAGGCGCG CTTTCGCGCG CAGCGCAGG GCATACGCGC
15 39541 CCGGAGCGCG TGCGTGGGCG GCGCTCCGCG GCGAGCGCGC CCGAGGTGGT CTCGTGTTT
39601 CCGGCGCAGG GCTCGCACTG GGTGGGCGAT GCGCGAAAGC TCATGGCCGA AGAGCCGGTC
15 39661 TTCCGGGCGG CGCTGGAGGG TTGCGACCGG GCCATCGAG CCGAAGCGGG CTGCTCGCTG
39721 CTCCGGGAGC TCTCCGCCGA CGAGGCGCGC TCGCAGCTCG GCGCATCGA CGTGGTTTAC
39781 CCGGTGCTCT TCGCCATGGA AGTAGCGCTT TCTGCGCTGT GCGGTCTGT GGGAGTGGAG
20 39841 CCGGAAGCGG TGGTGGGCCA CAGCATGGGC GAGGTGGCGC CCGGCGACGT GCGGCGCGC
39901 CTGTGCTCG AGSACGCGGT GCGGATCATC TGCCGGCGCA GCGGCTGCT GCGGCGGATC
20 39961 AGCGGTCCAG GCGAGATGGC GCTGGTCCAG CTGCTGCTGG AGGAGGCGCA GCGGCGGCTG
40021 CGTGGCCATG AGGGTCCGCT GAGCGTGGCG CTGAGCAACA GCGGCGCTC GACCGTGTCT
40081 GCGGCGCGC GCGGCGCGCT CTCGGAGGTG CTGGCGCGC TGACGGCCAA GGGGTGTTT
40141 TGGCGGCGAG TGAAGGTGGA CGTCCGCGC CATAGCCGC AGCTCGACCC GCTGCGCGAA
25 40201 GAGCTGATCG CCGGCTGGG GCGGATCCGG CCGGAGCGG CTGCGGTGCC GATGCGCTCG
40261 ACGGTGACGG CCGGGGTGAT CCGGGTCCG GAGCTCGGTG CAGCTACTG GCGGACAAAT
25 40321 CTTCGCGAGC CGGTGCGCTT CGCTGCGCGC GCGCAAGCGC TGCTGGAAG TGGCCCCACG
40381 CTGTTTCATG AGATGAGCCC GCACCCGATC CTGCTGCGCG CCCTGGACGA GATCCAGACG
40441 GCGSTCGAGC AAGGGGGCGC TGCGGTGGGC TCGCTGCGCG GAGGCGAGGA CAGCGCGCG
30 40501 ACGCTGCTGG AGGCGCTGGG GACGCTGTGG GCGTCCGGCT ATCCGTGAG CTGGGCTCGG
40561 CTGTTCCCG CCGGCGGCG AGGCGGTTCC CTGCGGACCT ATCCGTGAG GCACGAGCGG
40621 TGCTGATCG AGGTGAGCC TGACGCCCGC CGCTCGCGC CAGCGGACCC CACCAAGGAC
30 40681 TGGTCTACC GCGGACTG GCGGAGGTG CCGGCGCGC CCGGAAATC GGAGAACAGT
40741 CATGCGAGCT GGCTGCTGTT GCGGAGAGG GGTGGGTCG GCGAGGCGGT CGCTGCGAGC
35 40801 CTGTGAGCG GCGGACTTTC CTGCAACCTG CTTTCATGCG CTGCTGAGCG CTCCACCGTC
40861 GCGGAGCAGG TATCCGAAG TGCCAGTCCG CGAAACGACT GCGAGGAGT CCTCTACCTG
40921 TGGGCGCTCG ACGCGCTCGT CGATGCTGGG GCATCGGCGC ACGAAGTCAG CGAGGCTACC
40981 CCGCGTGCCA CCGCACCCGT CCTTGGGCTG GTTCGATTCC TGAGCGCTGC ECCCCATCCT
35 41041 CCTCGCTTCT GGGTGGTGAC CCGCGGGCA TGACGCTGG GCGGCGAGCC AGAGGTCTCT
41101 CTTTGCCAAG CCGGCTTGTG GCGCTCGCG CCGCTCGTGG CGCTGGAACA TCCGCTGCG
41161 TGGGGTGGCC TCGTGGACCT GGATCCTCAG AAGAGCCCGA CCGGAGTCGA GCGGCTGGT
41221 GCGGAGCTGC TTTCCCGGA CCGGAGGAT CACTGGCGT TCCGAGCGG TCGCGGCGC
41281 GCGGAGCGC TTGTAGCCGC CCGGCGGAG GCGGAGCTCG CACCGATATC GCTGTCCCG
41341 GAGGGAAGCT ACCTGGTGAC GGGTGGCTG GGTGGCTTG GTCTGCTCGT GGCTCGGTGG
45 41401 CTGGTGGAGC GGGGAGCTCG ACATCTGGTG CTACCGAGCC GCGGCGGCT GCGAGCGGA
41461 CAGGCGTCG GCGGAGAGCA GCGGCGGAG GCGGCGCGC GCATGCGAGC GGTGAGGGG
40 41521 CTGGAAGCAG AGGGCGCGC GGTGACCGTG GCGGCGGTGG ATGTCGCGA GGCCSATCCC
41581 ATGACGCGCG TCGTGGCCCG CATCGAGCC CCGTTGCGCG GGGTGGTGCA CGCGCCCGC
41641 GTCTTCCCG TGCGTCCCT GCGGAGAGC GAGGAGGCCC TGCTGAGTC GGTGCTCCGT
50 41701 CCGAAGGTGG CCGGAGCTG GCTGCTGAC CCGCTGCTGC GCGACCGGCC TCTCGACCTG
41761 TCTGTGCTGT TCTCGTCGGG CCGGCGGCTG TGGGTGGCA AAGGCCAAG CCGATACGCC
41821 GCGGCAATG CTTCTCTGA CCGGCTCGCG CACCATCGCC GCGGCGACTC CTGCGGCGC
45 41881 TTGAGCTCG CCTGGGCGCT ATGGGCGGAG GGAGGCGTG TTGATGCAA GGCTCATGCA
41941 CGTCTGAGC ACATCGGAG CTGCGCATG GCCACGGGC CCGCTTGTG GCGCTGGAG
55 42001 CCGCTGGTGA ACACGAGCG TGTCCAGCG TCGGTACAC GGAAGGACTG GCGGCGCTTC
42061 GCGGCGGTCT ATGCGCGCG AGGGCGGCG AACTTGCTTT CCGCTCTGGT CCGGAGGAC
42121 GAGCGCACTG CGTCTCCCC GGTGCGGAG CCAAACCGGA TCTGGCGCG CCGTCTCGTT
42181 GCGGAGAGCC GCTCAGCCCT CTACGAGCTG GTTCGCGCA TCGTCGCGC GGTGCTGGG
50 42241 TTCTCCGACC CCGGCGCGCT CGACGTCGCG CAGGCTTCG CCGAGCAGG GCTCGACTCC

5 42301 CTGATGGCTC TGGAGATCCG TAACCGCCTT CAGCGCGAGC TGGCGGAACG GCTGTCGGCG
42361 ACTCTGGCCT TCGACCACCC GACGGTGGAG CGGCTGGTGG CGCATCTCCT CACCGACGTG
42421 CTGPAAGCTGG AGGACCGGAG CGACACCCGG CACATCCGGT CGGTGGCGGC GGATGACGAC
42481 ATCGCCATCG TCGGTGCCGC CTGCCGGTTC CCGGGCGGGG ATGAGGGCCT GGAGACATAC
5 42541 TGGCGGCATC TGGCCGAGGG CATGGTGGTC AGCACCGAGG TGCCAGCCGA CCGGTGGCGC
42601 GCGCGGACT GGTACGACCC CGATCCGAG GTTCCGGGCC GGACCTATGT GGCCAAGGGG
10 42661 GCCTTCCTCC GCGATGTGCG CAGCTTGGAT GCGGCGITCT TCTCCATCTC CCTTCGTGAG
42721 GCGATGAGCC TGGACCCGCA ACAGCGGCTG TTGCTGGAGG TGAGCTGGGA GGCGATCGAG
42781 CGCGCTGGCC AGGACCCGAT GGCGCTGCGC GAGAGCGCCA CCGGCGTGTT CGTGGGCATG
10 42841 ATCGGGAGCG AGCACGCCGA GCGGGTGCAG GGCTTCGACG ACGACGCGGC GTTGCTGTAC
42901 GGCACCACCG GCAACCTGCT CAGCGTCGCC GCTGGACGGC TGTCCTTCTT CCTGGGTCTG
42961 CACGGCCCGA CGATGACGGT GGACACCGCG TGCTCGTCGT CGCTGGTGGC GTTGCACTC
15 43021 GCCTGCCAGA GCCTGCGATT GGGCGAGTGC GACCAGGCAC TGGCCGGCGG GTCCAGCGTG
43081 CTTTTGTGCG CGCGGTGCTT CGTCGCGGCA TCAGCGATGC GTTTGCTTTC GCCAGATGGG
15 43141 CCGTGCAAGA CGTTCGCGC GCGTGCAGAC GGCTTTGCGC GGGCCGAGGG CTGCGCCGTG
43201 GTGGTGTCTA AGCGGCTCCG TGACGCGCAG CGCCACCGCG ACCCATCCTT GGCGGTGGTC
43261 CGGAGCACGG CGATCAACCA CGATGGCCCG AGCAGCGGGC TCACGGTGCC CAGCGTCTCT
43321 GCGGAGCAGG CGTTGCTAGG CCAGGCGCTG GCGCAAGCGG GCGTGGCACC GGCCGAGGTC
43381 GATTTCTGTG AGTCCACCG GACGGGGACA GCGCTGGGTG ACCCGATCGA GGTGCAGCCG
20 43441 CTGGGCGCGG TGTATGGCCG GGGCCGCCCC GCGGAGCGGC CGCTCTGGCT GGGCGCTGTC
43501 AAGGCCAACC TCGGCCACCT GGAGGCCGCG GCGGCTTGG CCGGCGTGCT CAAGGTGCTC
43561 TTGGCGCTGG AGCACGAGCA GATTCCGGCT CAACCGGAGC TCGACGAGCT CAACCCGCA
43621 ATCCCGTGGG CAGAGCTGCC AGTGGCGCTT GTCCGCGCGG CGGTCCCCTG GCGCGCGGC
43681 GCGCGCCCGC GTCGTGACGG CCGTGGCGCT TTCGGCCTGA GCGGGACCAA CGCSCATGTG
25 43741 GTGTTGGAGG AGGCGCCGCG GGTGGAGCCT GAGSCCGCGG CCGGCGAGCG CGCTGCGGAG
43801 CTGTTCTGTC TGTCCGCGAA GAGCGTGGCG GCGCTGGATG CCGAGGCAGC CCGGCTGCGG
25 43861 GATCATCTGG AGAAGCATGT CGAGCTTGGC CTCGCGGATG TGGCGTTTCA CCTGGCGAGC
43921 ACGCGCAGCG CGATGGAGCA CCGGCTGGCG GTGGCCGCGA GCTCGCGCGA GGCGCTGCGA
43981 GGGGCGCTTT CCGCCGACAG GCAGGGGCAT ACGCCGCGCG GAGCCGTGCG TGGCGGGCC
30 44041 TCCGGCGGCA GCGCGCCGAA GGTGGTCTTC GTGTTTCCCG GCCAGGGCTC GCAGTGGGTG
44101 GGCATGGGCC GAAAGCTCAT GGCCGAAGAG CCGGTCTTCC GGGCGGCGCT GGAGGGTTGC
44161 GACCGGGCCA TCGAGGCGGA AGCGGGCTGG TCGCTGCTCG GGGAGCTCTC CGCCGACGAG
30 44221 GCGGCTCTCG AGCTCGGGCG CATCGACSTG GTTCAGCCCG TGCTCTTCGC CGTGGAGTA
44281 GCGCTTTCAG CGCTGTGGCG GTCGTGGGGA GTGGAGCCGG AAGCGGTGGT GGGCCACAGC
35 44341 ATGGGCGAGG TTGC3GCGCG GCACGTGGCC GCGCGCTGT CGCTCGAGGA TGCGGTGGCG
44401 ATCATCTGCC GCGCGAGCCG GCTGCTGCGG CGGATCAGCG GTCAGGGCGA GATGGCGCTG
44461 GTCGAGCTGT CGCTGGAGGA GGCCGAGGCG GCGCTGCGTG GCCATGAGGG TCGGCTGAGC
44521 GTGGCGGTGA GCAACAGCCC GCGCTCGACC GTGCTCGCAG GCGAGCCGCG GGCGCTCTCG
44581 GAGGTGCTGG CCGCGCTGAC GGCCAGGGGG GTCTTCTGGC GGCAGGTGAA GGTGGACGTC
35 44641 GCCAGCCATA GCGCGCAGGT CGACCCGCTG CGCGAAGAGC TGGTTCGCGC GCTGGGAGCG
44701 ATCCGGCCCG GAGCGGCTGC GGTGCCGATG CGCTCGACCG TGACGGGCGG GGTGATTGCG
44761 GSTCCGAGC TCGGTGCGAG CTACTGGGCG GACAATCTTC GGCAGCCGGT GCGCTTCGCT
44821 GCGGCGGCGC AAGCGCTGCT GGAAGGTGCG CCCACGCTGT TCATCGAGAT GAGCCCGCAC
44881 CCGATCCTGG TSCCGCCTCT GGACGAGATC CAGACGGCGG TCGAGCAAGG GGGCGCTGCG
45 44941 GTGGGCTCGC TCGGGCGAGG GCAGGACGAG CGCGCGACCG TGCTGGAGGC GCTGGGGACG
45001 CTGTGGGCGT CCGGCTATCC GGTGAGCTGG GCTCGGCTGT TCCCCGCGGG CCGCAGGCGG
45061 GTTCCGCTGC CGACCTATCC CTGGCAGCAC GAGCGGTACT GGATCGAGGA CAGCGTGCTA
45121 GGGTCGAAGC CCTCGCTGCG GCTTCGGCAG CTTCTATAAC GCGCCACGGA CCATCCGCTG
45181 CTCGGGGCTC CATTGCTGCT CTCGCGCGCA CCGGAGGCTC ACTTGTGGGA GCAAGCGCTG
50 45241 AGCGACGAGA GGCTATCCTA TCTTTCGGAA CATAGGGTCC ATGGCGAAGC CGTGTGGCCC
45301 AGCGCGGCGT ATGTAGAGAT GGCGCTCGCC GCGGGCGTAG ATCTCTATGG CCGGGCGAGC
45361 CTGGTGCTGG AGCAGCTGGC GCTCGAGCGA GCCCTCGCCG TGCTTCCGA AGGCGGACGC
45421 ATCGTGCAAG TGGCCCTCAG CGAAGAAGGG CCGGTCGGG CCTATTCCA GGTATCGAGC
45481 CGTGAGGAGG CAGGTAGAAG CTGGGTTCGG CACGCCACGG GGCACGTGTG TAGCGACCAG
55 45541 AGCTCAGCAG TGGGAGCGTT GAAGGAAGCT CCGTGGGAGA TTCAACAGCG ATGTCCGAGC
45601 GTCCTGTCTG CCGAGGCGCT CTATCCGCTG CTCACGACC ACGCCCTCGA CTATGCCCCC
45661 TCCTTCCAGG GTGTGAGCA GGTGTGGCTC GGCACGGGGG AGGTGCTCGG CCGGTACGCG
45721 TTGCCAGAAG ACATGGCATC CTCAAGTGGC GCTATCGGA TTCATCCCGC CTTGTGGAT
50 45781 GCATGTTTTT AAGTGCTGAC CGCGCTGCTC ACCACGCGG AATCCATCGA GATTCGGAGG

5 45841 CGGCTGACGG ATCTCCACGA ACCGGATCTC CCGCGGTCCA GGGCTCCGGT GAATCAAGCG
45901 GTGAGTGACA CCTGGCTGTG GGACSCCGCG CTGGACGCTG GACGGCGCCA GAGCGCGAGC
45961 GTGCCCCTCG ACCTGGTGTG CCGCAGCTTC CACGCGAAGT GGGAGGTCAT SGATCGCCTC
5 46021 GCGCAGACGT ACATCATCCG CACTCTCCGC ACATGGAACG TCTTCTGCGC TGCTGGAGAG
46081 CGTCACACGA TAGACGAGTT GCTCGTCAGG CTCCAAATCT CTGCTGTCTA CAGGAAGGTC
46141 ATCAAGCGAT GGATGGATCA CCTTGTGCGG ATCGGCGTCC TTGTAGGGGA CGGAGAGCAT
10 46201 CTTGTGAGCT CTCAGCCGCT GCCGGAGCAT GATTGGGCGG CCGTGTCTGA GGAGGCCGCG
46261 ACGGTGTTCG CCGACCTCCC AGTCCTACTT GAGTSGTGCA AGTTTGCCGG GGAACGGCTC
46321 GCGGACGTGT TGACCGGGAA GACGCTGGCG CTCGAGATCC TCTTCCCTGG CCGCTCGTTC
10 46381 GATATGGCGG AGCGAATCTA TCAAGATTCTG CCCATCGCCC GTTACTCGAA CGGCATCGTG
46441 CGCGGTGTCT TCGAGTCGGC GGCGCGGGTG GTAGCACCGT CCGGAACGTT CAGCATCTTG
46501 GAGATCGGAG CAGGGACGGG CGCGACCACC GCCGCCGTTC TCCCGGTGTT GCTGCCGTGAC
15 46561 CGGACAGAAT AUCATTTTAC CGATGTTTCT CCGCTCTTCC TTGCTCGTGC GGAGCPAAGA
46621 TTTCGAGATC ATCCATTCTT GAAGTATGGT ATTCTEGATA TCGACCAGGA GCCAGCTGGC
15 46681 CAGGGATACG CACATCAGAA GTTCGACGTC ATCGTCGCGG CCAACGTCAT CCATGCGACC
46741 CGCGATATAA GAGCCACCGC GAAGCGTCTC CTGTCGTTCG TCGCGCCCGG AGGCCCTTCTG
46801 GTGCTGGTCTG AGGGCACAGG GCATCCGATC TGGTTCGATA TCACCACGGG ATTGATCGAG
46861 GGGTGGCAGA AGTACGAAGA TGATCTTCGT ACCGACCATC CGCTCCTGCC TCTCGGACC
20 46921 TGSTGTGACG TCTGCGCCG GGTAGSCTTT GCGGATGCCG TGAGTCTGCC AGGCGACGGA
46981 TCTCCGGCGG GGATCCTCGG ACAGCACTG ATCCTCTCGC GCCCTCCGGG CATAGCAGGA
47041 GCGCTTGTG ACAGCTCCGG TGAGTCGGCG ACCGAATCGC CCGCCGCGCG TGCACTACGG
47101 CAGGAATGGG CCGATGGCTC CGCTGACGGC GTCCATCGGA TGGCGTTGGA GAGAATGTAC
47161 TTCCACCGCC GGCCGGGCCG GCAGGTTTGG GTCCACCGTC GATTGCGTAC CGGTGGAGGC
47221 GCGTTCACGA AGGCGCTCAC TGEAGATCTG CTCCTGTTCG AAGAGACCGG GCAGGTCGTG
25 47281 GCAGAGGTTT AGGGGCTCCG CCTGCCCGAG CTCGAGGCTT CTGCTTTCGC GCCCGGGAC
47341 CCGCGGGAAG AGTGGTTGTA CGCCTTGGAA TGGCAGCGCA AAGACCCTAT ACCAGAGGCT
25 47401 CCGGCAGCCG CGTCTTCTTC CACCGCGGGG GCTTGGCTCG TGCTGATGGA CCAGGGCGGG
47461 ACAGGCGCTG CGCTCGTATC GCTGCTGGAA GGGCGAGCG AGGCGTGCGT GCGCGTCGTG
47521 GCGGTACGG CATAAGCCTG CCTCGCGCCG GGGCTGTATC AAGTCGATCC GCGCGAGCCA
30 47581 GATGGCTTTC ATACCCTGCT CCGCGATGCA TTCGGCGAGG ACCGGATGTG CCGCGCGGTA
47641 GTGCATATGT GGAGCCTTGA TGCGAAGGCA GCAGGGGAGA GGACGACAGC GGAGTCECTT
47701 CAGGCCGATC AACTCCTGGG GAGCCTGAGC GCGCTTTCG TGGTGCAGGC GCTGGTGGCG
30 47761 CCGAGGTGGC GCAACATGCC GCGACTTTGG CTCTTGACCC GCGCCGTGCA TCGGTTGGG
47821 GCGSAGGACG CAGCGGCCTC GGTGGCGCAG GCGCCGGTGT GGGGCCTCGG TCGGACGCTC
35 47881 GCGCTCGAGC ATCCAGAGCT GCGGTGCACG CTCGTGGACG TGAACCCGGC GCGCTCTCCA
47941 GAGGACGCAAG CTGCACTCGC GGTGGAGCTC GGGCGAGCG ACAGAGAGGA CCAGATCGCA
48001 TTGCGCTCGA ATGGCCGCTA CGTGGCGCGC CTCGTGCGGA GCTCCTTTC CGGCAAGCCT
48061 GCTACGGATT GCGGCATCCG GCGGACCGC AGTTATGTGA TCACCGATGG CATGGGAGA
35 48121 GTGGGGCTCT CCGTCCGCGA ATGGATGGTG ATGCAGGGG CCGCCATGT GGTGCTCGTG
48181 GATCGCGGCG GCGCTTCCGA CGCCTCCCGG GATGCCCTTC GGTCCATGGC CGAGGCTGGC
48241 GCAGAGGTGC AGATCGTGGA GGCCGACGTG GCTCGGCGCG TCGATGTGCG TCGGCTTCTC
48301 TCGAAGATCG AACCGTCGAT GCGCGCGCTT CCGGGGATCG TGTACGTGGA CCGGACCTTC
48361 CAGGGCGACT CCTCGATGCT GGAGCTGATG GCGCATCGCT TCAAGGAGTG GATGTATCCC
48421 AAGTGCTCG GAGCGTGCAA CCGGACCGG CTGACCAGGG ATAGATCGCT GGACTTCTTC
45 48481 GTCCTGTAAT CCTCGGGCAC CTCGCTTCTG GCGTTCGCCG GACAGGGGAG CCGCGCCGCC
48541 GGTGACGCTT TCTTGGACGC CATCGCGCAT CACCGGTGTA GGCTGGGCCT CACAGCGATG
48601 AGCATCAACT GGGGATTGCT CTCCGAAGCA TCATCGCGCG CGACCCCSAA CGACGGCGGC
48661 GCACGGCTCC AATACCGGGG GATGGAAGGT CTCACGCTGG AGCAGGGAGC GGAGGCGCTC
48721 GGGCGCTTCC TCGCACAACC CAGGSCGCGG GTAGGGGTAA TCGGCTGAA TCTGCGCCAG
50 48781 TGGCTGGAGT TCTATCCCAA CGCGSCCGGA CTGGCGCTGT GGGCGGAGTT GCTGAAGGAG
48841 CGTGACCGCA CCGACCGGAG CGCGTCGAAC GCATCGAACC TCGCGAGGC GCTGCAGAGC
48901 GCCAGGCCCG AAGATCGTCA GTTGGTCTTG GAGAAGCACT TGAGCGAGCT GTTGGGGCGG
48961 GGGCTGCGCC TTCCGCGCGA GAGGATCGAG CCGCACGTGC CGTTCAGCAA TCTCGGCATG
45 49021 GACTCGTTGA TAGGCCTGGA GCTCCGCAAC CGCATCGAGG CCGCGCTCGG CATCACCGTG
55 49081 CCGGCGACCC TGCTATGGAC TTACCCTACC GTAGCAGCTC TGAGCGGGAA CCTGCTAGAT
49141 ATTCTGTTCC CGAATGCCGG CCGGACTCAC GCTCCGGCCA CCGA3CGGGA GAAGAGCTTC
49201 GAGAACGATG CCGCAGATCT CGAGGCTCTG CCGGGTATGA CCGACGAGCA GAAGGACCGG
49261 TTGCTCGCCG AAAAGCTGGC GCAGCTCGCG CAGATCGTTG GTGAGTAAGG GACTGAGGGA
50 49321 GTATGSCGAC CACGAATGCC GGAAGCTTG AGCATGCCCT TCTGCTCATG GACAAGCTTG

5 49381 CGAAAAAGAA CGCGTCTTTG GAGCAAGAGC GGACCGAGCC GATCGCCATC ATAGGTATTG
49441 GCTGCCGCTT CCCCAGCGGA GCGGACACTC CGGAGGCATT CTGGGAGCTG CTCGACTCGG
49501 GCCGAGACGC GGTCCA3CCG CTCGACCGGC CCTGGGCGCT GGTCGGCGTC CATCCCAGCG
5 49561 AGGAGGTGCC GCGCTGGGCC GGACTGCTCA CCGAGGCGGT GGACGGCTTC GACGCCGCTT
49621 TCTTTGGCAC CTCGCTCGG GAGCGCGGGT CGCTCGATCC TCAGCAACGC CTGCTGCTGG
49681 AGGTACCTG GGAAGGGCTC GAGGACGCG GCATCGCACC CCAGTCCCTC GACGGCAGCC
10 49741 GCACCGGGT ATTCTTGGG GCATGCAGCA GCGACTACTC GCATACCGTT GCGCAACAGC
49801 GGCGCGAGGA GCAGGACGCG TACGACATCA CCGGCAATAC GCTCAGCGTC GCCGCCGGAC
49861 GGTGTCTTA TACGCTAGGG CTGACGGGAC CCTGCCTGAC CGTCGACACG GCCTGCTCGT
10 49921 CGTCGCTCGT GGCCATCCAC CTTGCTGCG GCAGCCTGCG CGCTCGCGAG AGCGATCTCG
49981 CGCTGGCGGG GGGCGTCAAC ATGCTCCTTT CGTCCAAGAC GATGATAATG CTGGGGCGCA
50041 TCCAGGCGCT GTCGCCCCGAT GGCCACTGCC GGACATTGCA CGCCTCGGCC AACGGGTTG
15 50101 TCCGTGGGGA GGCTTGGCGT ATGGTCTGTC TCAAACGGCT CTCCGACGCC CAGCGACATG
50161 GCGATCGGAT CTGGGCTCTG ATCCGGGTT CGGCCATGAA TCAGGATGGC CGGTCCAGAC
15 50221 GGTGATGGC ACCCAATGTG CTCGCTCAGG AGGCGCTCTT ACGCCAGGCG CTGCAGAGCG
50281 CTCGCGTGA CGCCCGCGGC ATCGATTATG TCGAGACCCA CGGAACGGGG ACCTCGCTCG
50341 GCGACCCGAT CGAGGTCGAT GCGCTGCGTG CCGTGATGGG GCCGGCGCGG GCCGATGGGA
50401 GCCGCTGCGT GCTGGGCGCA GTGAAGACCA ACCTCGGCGA CCTGGAGGEC GCTGCAGGCG
20 50461 TGGCGGGTTT GATCAAGGCG GCGCTGGCTC TGCACCAAGA ATCGATCCCG CGAAACCTCC
50521 ATTTTCACAC GCTCAATCCG CGGATCCGGA TCGAGGGGAC CGCGCTCGCG CTGGCGACGG
20 50581 AGCCGGTGCC GTGGCGCGCG GCGGGCCGAC CGCGCTTCGC GGGGGTGAGC GCGTTCGGCC
50641 TCAGCGGCAC CAACGTCCAT GTCGTGCTGG AGGAGGCGCC GGCCACGGTG CTCGCACCGG
50701 CGACGCGGG GCGCTCAGCA GAGCTTTTCG TGCTGTGCGC GAAGAGCACC GCGCGCTGG
25 50761 ACGCACAGCC GCGCGGGCTC TCAGCGCACA TCGCCGCGTA CCUGGAGCAG GGCTCGGAG
50821 ACGTCGCGTT CAGCCTGGTA GCGACGCGGA GCCCGATGGA GCACCGGCTC GCGGTGGCGG
25 50881 CGACCTCGCG CGAGGCGCTG CGAAGCGCGC TGAAGCTGC GCGCGAGGGG CAGAUCCCGG
50941 CAGGCSGCGC GCGCGGCGG GCGCTTCTT CGCCCGGCAA GCTCGCCTTC CTGTTCCCGG
51001 GGCAGSGCGC GCAGGTGCGG GGCATGGGCG GTGGGTTGTG GAGGCGTGG CCGGCGTTCC
51061 GCGAGACCTT CGACCGGTGC GTCACGCTCT TCGACCGGGA GCTCCATCAG CCGCTCTGCG
30 51121 AGGTGATGTG GGCCGAGCCG GGCAGCAGCA GGTGCTCGTT CCTGGACCAG ACGGCATTCA
51181 CCGAGCCGGC GCTCTTTGCG CTGGASTACG CGCTGGCCGC GCTCTTCCGG TCGTGGGGCG
51241 TGGAGCCGGA GCTCATCGCT GGCCATAGCC TCGGCGAGCT GGTGGCCGCG TCGTGGGGCG
30 51301 GTGTGTTCTC CCTCGAGGAC GCCGTGCGCT TGCTGGTCCG GCGCGGCCG3 TTGATGCAGG
51361 CGCTGCCGGC CGGCGGTGCG ATGGTATCGA TCGCCGCGCC GGAGGCCGAC GTGGCTGCCG
35 51421 CGGTGGCGCC GCACGCAGCG TC3GTGTGA TCGCGGCGAG CAATGGGCCG GAGCAGGTGG
51481 TGATCGCGGG CGCCGAGAAA TTCGTGCAGC AGATCGCGGC GCGGTTTCGCG GCGCGGGGGG
51541 CGCGAACCAA ACCGCTGCAT GTTTCGCAGC CGTTCCACTC GCCGCTCATG GATCCGATGC
51601 TGGAGGCGTT CCGGCGGGTG ACCGAGTCGG TGACGTATCG GCGGCCCTTCG ATGGCGCTGG
35 51661 TGAGCAACCT GAGCCGGAAG CCCTGCACGG ATGAGGTGTG CCGCGCCGGT TACTGGGTGC
51721 GTCACGCGCG AGAGGCGGTG CGCTTCGCGG ACGCGGTGAA GCGCGTGCAC GCGGCC3GTG
51781 CCGGCATCTT CGTCGAGGTG GGCCCGAAGC CCGCGCTGCT CCGCCTTTTG CCGCCTGCC
51841 TGCCGGATGC CAGGCCGGTG CTGCTCCAG CGTCGCGCGC CCGGCGTGAC GAGGCTGCCA
51901 GCGCGCTGGA GCGCTGGGT GGGTCTGCGG TCGTGGGTGG ATCGGTCACC TGGTGGGTG
45 51961 TCTTCCCTTC GGGCGGACGG CCGGTACCGT TGCCAACCTA TCCCTGGCAG CGCGAGCGTT
52021 ACTGGATCGA AGCGCCGGTC GATGGTGAGG CGGACGGCAT CCGCCGTGCT CAGGCGGGGG
40 52081 ACCACCCCT TCTGGGTGAA GCCTTTTCCG TGTCGACCCA TGCCGGTCTG CGCCTGTGGG
52141 AGACGACGCT GGACCGAAAG CCGCTGCCGT GGCTCGGCGA GCACCGGGCG CAGCGGGAGG
52201 TCGTGTTCCT TGGCGCCGGG TACCTGGAGA TGGCGCTGTC GTCGGGGGCC GAGATCTTGG
52261 GCGATGGACC GATCCAGGTC ACGGATGTGG TGCTCATCGA GACGCTGACC TTCGCGGGCG
50 52321 ATACGGCGGT ACCGCTCCAG GTGGTGACGA CCGAGGAGCG ACCGGGACGG CTGCGGTTCC
52381 AGGTAGCGAG TCGGGAGCCG GGGGCACGTC GCGCTCCTT CCGGATCCAC GCCCGCGCG
45 52441 TGCTGCGCCG GGTGCGGCGC GCGGAGACCC CCGCGAGGTT GAACCTCGCC GCCCTGCGCG
52501 CCCGCTTCA TGCCGCGGTG CCCGCTGCGG CTATCTATGG GCGGCTCGCC GAGATGGGGC
52561 TTCAATACGG CCCGCGGTTG CCGGSGCTCG CCGAGCTGTG GCGGGGTGAG GCGGAGGCGC
55 52621 TGGGCGAGGT GAGACTGCCT GATCCGCGG GCTCCGCGAC AGCCTACCAG CTGCATCCCG
52681 TGCTGCTGGA CCGGTGCGTC CAAATGATTG TTGGCGCGTT CGCCGATCGC GATGAGGCGA
52741 CGCCGTGGGC GCGGTGGAG GTGGGCTCGG TCGGCTGTT CAGCGGTCT CCTGGGGAGC
52801 TATGGTGCCA TGCGCGCGTC GTGAGCGATG GTCAACAGGC CCCAGCCGG TGGAGCGCCG
50 52861 ACTTTGACTT GATGGACGST ACGGGCGCGG TGCTCGCCGA GATCTCCCG CTGGTGTGG

5 52921 AGCGGCTTGC GAGCGGTGTA CGCCGGCGCG ACGCAGACGA CTGGTTCCCTG GAGCTGGATT
52981 GGGAGCCCCG GCGGCTCGAG GGGCCCAAGA TCACAGCCGG CCGGTGGCTG CTGCTCGGCG
53041 AGGGTGGTGG GCTCGGGCGC TCGTTGTGCT CAGCGCTGAA GGCCGCCGGC CATGTCGTCC
5 53101 TCCACGCCGC GGGGGACGAC ACCAGCGCTG CAGGAATGCG CGCGCTCCTG GCCAACGCGT
53161 TCGACGGCCA GGCCCGACG GCGGTGGTGC ACCTCAGCAG CCTCGACGGG GCGGGCCAGC
53221 TCGACCCGGG GCTCGGGCGC CAGGGCGGCG TCGACGCGCC CCGGAGCCCA GATGTCGATG
10 53281 CCGATGCCCT CGAGTCGGCG CTGATGCGTG GTTGGCAGAG CGTGCTCTCC CTGGTGCAAG
53341 CGCTGGTCCG CATGGACCTC CGAAATGCGC CGCGGCTGTG GCTTTTGACC CGCGGGGCTC
53401 ACGCGGCCGC CGCCGGCGAT GTCTCCGTGG TGCAAGCGCC GCTGTTGGGG CTGGGCCGCA
10 53461 CCATCGCCTT GGAGCAGGCC GAGCTGCGCT GTATCAGCGT CGACCTCGAT CCAGCCACAGC
53521 CTGAAGGGGA AGCCGATGCT TTGCTGGCCG AGCTACTTGC AGATGATGCC GAGGAGGAGG
53581 TCGCCCTGCG CGGTGGCGAG CCGTTTGTGG CGCGGCTCGT CCACCGGCTG CCCGAGGCTC
15 53641 AACGCCGGGA GAAGATCGCG CCCGCCGGTG ACAGGCCGTT CCGGCTAGAG ATCGATGAAC
53701 CCGGCGTGCT GGACCAACTG GTGCTCCGGG CCACGGGGCG GCGCGCTCCT GGTCCGGGCG
15 53761 AGGTCGAGAT CGCCGTCGAA GCGGCGGGG TCGACTCCAT CGACATCCAG CTGGCGGTGG
53821 GCGTTGCTCC CAATGACCTG CCTGGAGGAG AAATCGAGCC GTCGGTGCTC GGAAGCGAGT
53881 GCGCCGGGCG CATCGTCGCT GTGGGCGAGG GCGTGAACGG CCTTGIGGTG GGCCAGCCCG
53941 TGATCGCCCT TCGGCGGGGA GTATTTGCTA CCCATGTCAC CACGTCGGCC ACGCTGGTGT
20 54001 TGCCTCGGCC TCTGGGGCTC TCGGCGACCG AGGCGGCCGC GATGCCCTC GCGTATTTGA
54061 CGCCCTGGTA CGCCCTCGAC AAGGTCGCCC ACCTGCAGGC GGGGGAGCCG GTGCTGATCC
54121 GTGCGGAGGC CCGTGGTATC GGTCTTTGCG CCGTGCGATG GCGCGAGCGC GTGGGCGCCG
54181 AGGTGTATGC GACCGCCGAC ACGCCCGAGA AACGTGCCTA CCTGGAGTCG CTGGGCGTGC
54241 GGTACGTGAG CGATTCCCGC TCGGGCCGGT TCGCCGAGA CCTGCATGCA TGGACGGACG
25 54301 GCGAGSGTGT GGACGTCGTG CTCGACTCGC TTTCGGGCGA GCACATCGAC AAGAGCCTCA
54361 TGGTCTGCG CGCCTGTGGC CGCCTGTGTA AGCTGGGCG GCGCGACGAC TCGCCGACA
54421 CGCAGCCTGG GCTGCCGCGC CTCCTACGGA ATTTTTCCTT CTCGAGGTG GACTTGGCGG
25 54481 GAATGATGCT CGATCAACCG GCGAGGATCC GTGCGCTCCT CGACGAGCTG TTCGGGTGG
54541 TCGCAGCCGG TGCCATCAGC CCACTGGGGT CCGGGTTGCG CGTTGGCGGA TCCCTCACGC
54601 CACCGCCGGT CGAGACCTTC CCGATCTCTC GCGCAGCCGA GGCATTCCGG AGGATGGCGC
30 54661 AAGGACAGCA TCTCGGGAAG CTCGTGCTCA CGCTGGACGA CCCGGAGGTG CCGATCCGCG
54721 CTCCGGCCGA ATCCAGCGTC GCGGTCCGCG CCGACGGCAC CTACCTGTG ACCGGCGGTC
54781 TGGGTGGGCT CCGTCTGCGC GTGGCCGGAT GGCTGGCCGA GCGGGGCGCG GGGCAACTGG
30 54841 TGCTGGTGGG CCGTCCGGT GCGGCGAGCG CAGAGCAGCG AGCCGCCCTG GCGGCGCTAG
54901 AGGCCCCAGG CCGCGCGGTC ACGGTGGCGA AAGCGGATGT CGCCGATCGG TCACAGATCG
35 54961 AGCGGTCCT CCGCGAGGTT ACCGCTCGG GGATGCCGCT GCGGGGTGTC GTGCATGCGG
55021 CAGGTCTTGT GGATGACGGG CTGCTGATGC AGCAGACTCC GGCGCGGCTC CGCACGGTGA
55081 TGGGACCTAA GGTCCAGSGA GCCTTGCACT TGCACAGCT GACACGCGAA GCGCTCTTT
55141 CTTCTTCTGT GCTGTACGCT TCTGCAGCTG GGCTGTTCCG CTCGCCAGGC CAGGGCAACT
35 55201 ATGCCGCGAG CAACGCGTTC CTCGACGCCC TTTCGCATCA CCGCAGGGCG CACGGCCTGC
55261 CCGCGCTGAG CATCGACTGG GGCATGTTCA CCGAGGTGGG GATGGCCGTT GCGCAAGAAA
55321 ACCGTGGCGC GCGCTGATC TCTCGCGGGA TCGGGGCGAT CACCCCGAT GAGGGTCTGT
55381 CAGCTCTGGC GCGCTTGCTC GAGGGTGATC GCGTGACAG GGGGGTGATA CCGATCACTC
55441 CCGGCGAGTG GGTGGAGTTC TACCCGGCAA CAGCGGCCCTC ACGGAGGTTG TCGCGGCTGG
45 55501 TGACCACGCA GCGCGCGGTT GCTGATCGGA CCGCCGGGGA TCGGGACCTG CTCGAACAGC
55561 TTCCCTCGGC TGAGCCGAGC GCGCGGGCGG GGCTGCTGCA GGACGTCTG CCGGTGACG
40 55621 TCTCGCATGT GCTGCGTCTC CTGAAGACA AGATCGAGGT GGATGCCCCG CTCTCGAGCA
55681 TGGGCATGGA CTCGCTGATG AGCCTGGAGC TCGCAACCG CATCGAGGCT GCGCTGGGCG
55741 TCGCCGCGCC TCGAGCCTTG GGTGGAGCT ACCCAACGGT AGCAGCGATA ACGCGCTGEC
55801 TGCTCGACGA CGCCCTCGCC GTCCGGCTTG GCGGCGGGTC GGACACGGAC GAATCGACGG
50 55861 CAAGCSCCGG ATCGTTCTGC CACGTCTCTC GCTTTCTGTC TGTCTCAAG CCGCGGCTC
55921 GTCTCTCTG TTTTACGGT TCTGGCGGCT CCGCCGAGGG CTTCGTTCC TGGTCGGAZA
55981 AGTCTGAGTG GAGCGATCTG GAAATCGTGG CCATGTGGCA CGATCGCAGC CTCGCCTCCG
45 56041 AGGACGCGCC TGGTAAGAAG TACGTCCAAG AGGCGGCCCTC GCTGATTCAG CACTATGCAG
56101 ACGCACCGTT TGCGTTAGTA GGGTTCAGCC TGGGTGTCCG GTTCGTCATG GGGACAGCCG
55 56161 TGGAGCTCGC TAGTCGTTC GCGCACCGG CTCCGCTGGC CGTTTTGCG TTGGGCGGCA
56221 GCTTGATCTC TTCTTCAGAG ATCACCCTCG AGATGGAGAC CGATATAATA GCCAAGCTCT
56281 TCTTCCGAAA TGCCGCGGGT TCGTGCGAT CCACCAACA AGTTCAGGCC GATGCTCGCC
56341 CAGACAAGGT CATCACAGAC ACCATGGTGG CTCCGGCCCC CCGGACTCG AAGGAGCCCG
50 56401 CCTCGAAGAT CCGGTCCCT ATCGTCGCCA TCGCCGGCTC GGACGATGTC ATCGTGCCTC

5 56461 CAAGCGACGT TCAGGATCTA CAATCTCGCA CCACGGAGCG CTTCTATATG CATCTCCTTC
56521 CCGGAGATCA CGAGTTTCTC GTCGATCGAG GGCGCGAGAT CATGCACATC GTCGACTCGC
56581 ATCTCAATCC GCTGCTCGCC GCGAGGACGA CGTCGTACAG CCCCCTTTC GAGGCAAAAT
56641 GATGGCAGCC TCCCTCGGGC GCGCGAGATG GTTGGGAGCA GCGTGGGTGC TGGTGGCCGG
5 56701 CCGCAGGCAG CGGAGGCTCA TGAGCCTTCC TGGAGTTTG CAGCATAGGA GATTTTATGA
56761 CACAGGAGCA AGCGAATCAG AGTGAGACGA AGCCTGCTTT CCACTTCAAG CCGTTCGCGC
10 56821 CTGGGTACGC GGAGGACCCG TTTCCCGCGA TCGAGCGCCT GAGAGAGGCA ACCUCCATCT
56881 TCTACTGGGA IGAAGGCCGC TCCTGGGTCC TCACCCGATA CCACGACGTG TCGGCGGTGT
56941 TCCGCGACGA ACCTTCGCG GTCAGTCGAG AAGAATGGGA ATCGAGCGCG GAGTACTCGT
10 57001 CCGCCATTCC CGAGCTCAGC GATATGAAGA AGTACGGATT GTTCGGGCTG CCGCCCGAGG
57061 ATCAGCTCG GGTCCGCAAG CTCGTCAACC CATCGTTTAC GTCACGCGCG ATCGACCTGC
57121 TCGCGCGCGA AATACAGCCC ACCGTGCGAC AGCTGCTCGA TGCTCGCTCC GGACAAGAGG
57181 AGTTCGACGT TGTGCGGGAT TACGCGGAGG GAATCCCGAT GCGTGGCATC AGCGCTCTGT
15 57241 TGAAGGTTCC GGCCGAGTGT GACGAGAAGT TCCGTGCTT CCGCTCGGCG ACTGCGCGCG
57301 CGCTCGGCTT GGGTTTGGTG CCCCCTGCG ATGAGGAGAC CAAGACCTTG GTCGCTCCG
15 57361 TCACCGAGGG GCTCGCGCTG CTCCATGGCG TCCTCGATGA GCGGCGCAGG AACCCGCTCG
57421 AAAATGACGT CTTGACGATG CTGCTTCAGG CCGAGGCCGA CCGCAGCAGG CTGAGCACGA
57481 AGGAGCTGGT CCGCTCGTG GGTGCGATTA TCGCTGCTGG CACCGATACC ACGATCTACC
57541 TTATCGCGTT CGCTGTGCTC AACCTGCTGC GGTGCGCCGA GCGCTCGAG CTGGTGAAGG
20 57601 CCGAGCCCGC GTCATGAGG AACCGCTCG ATGAGGTGCT CCGCTTCGAC AATATCCTCA
57661 GAATAGGAAC TGTGCGTTTC GCCAGGCAGG ACCGAGGTA CTGCGGGGCA TCGATCAAGA
57721 AAGGGGAGAT GGTCTTTCTC CTGATCCCGA GCGCCCTGAG AGATGGGACT GTATTCTCCA
57781 GCCCAACGT GTTTGATGTG CGACGCGACA CGAGGCGGAG CCTCCCGTAC GGTAGAGGCC
57841 CCCATGCTG CCGCGGGTG TCCCTTGCTC CCCTCGAGGC GSAGATCGCC GTGGGCACCA
25 57901 TCTTCCGTAG GTTCCCGGAG ATGAAGCTGA AAGAACTCC CGTGTGGA TACCACCCCG
57961 CGTTCGGGAA CATCGAATCA CTCAACGTCA TCTTGAAGCC CTCCAAAGCT GSATAACTCG
25 58021 CGGGGGCATC GCTTCCCGAA CCTCATCTT TCATGATGCA ACTCGCGCGC GGTGCTGTC
58081 TCGCGCGGGT GCGATTGAT CCAGCGGACA AGCCCATGT CAGCGCGCGA AGATCGAATC
58141 CACGGCCCGG AGAAGAGCCC GATGGCGAGC CCGTCCGGGT AACGTCGGAA GAAGTGCCGG
30 58201 GCGCCGCCCT GGGAGCGCAA AGCTCGCTCG CTCGCGCTCA GCGCGCCGCT TGCCATGTCC
58261 GGCCCTGCAC CCGCACCAG GAGCCACCCG CCTGATGCA CCGCCTCACC GAGCGGCGAG
58321 TCTGCTCTC GCTCGTCCG CTGCGCTCG TCCTCCGAC CCGCGCGCGC TTCGCGGAGC
30 58381 TCGCGCGGGC GCTGCGCCAG CCGGAGGTG TCGGCGAGCT CTTGCGCGGC GTGGTGTG
58441 GCGCGTCCGT CGTCCGCGCG CTCGCTCCTG GGTTCATCG AGTCTCTTC CAGGATCCGG
35 58501 CGGTCCGGGG CGTGTCTCC GGCATCTCCT GGATAGGCG GCTCGTCTG CTGCTCATGG
58561 CCGGTATCGA GGTGATGTG AGCATCTAC GCAAGGAGG GCGCCCGGG GCGCTCTCGG
58621 CGCTCGGCGC GATCGCGCCC CCGCTGCGCA CCGCGGGCC GCTGGTGCAG CGCATGCAGG
58681 GCACGTTGAC GTGGGATCTC GACGTCTCG CCGCAGCTC TCGCAAGCC TGAGCCTCGG
35 58741 CGCCTGCTCG TACACCTCG CCGTGTCTCG TCCGCGCGCG GACATCCGGC CCGCCCGCGC
58801 GCGCCAGCTC GAGCCGACT CCGCGATGA CGAGGCCGAC GAGGCGCTCC GCGGTTCCG
58861 CGACGCGATC GCGCGTACT CCGAGGCCGT TCGGTGGGCG GAGGCGGCG AGCGGCCGCG
58921 GCTGGAGAGC CTCGTGCGGC TCGCATCGT GCGGTGGGCG AAGGCGCTCG ACAAGGCACC
58981 TTTGCGGCAC ACGACGCCG CCGTCTCCA GATCGCGGC AGACTTCCC AAAAAACGAA
45 59041 TCGGCTCTGG TTTGATGTC CCGCCCGGTA CCGGAGCTC CCGCGGCGA CCGGAGCAGC
59101 GCTCCGCGAC GCGGCTCGC CCACGGAGG GCTC3CGGC GGCCCGTACC GCGGATCGAG
40 59161 CAGCGTGTCC GCTGCCGTAG GGGAGTTTCG GGGGAGGCG GCGCCCTTC ACCCCGCGEA
59221 CCGCGTACCC GCGTCCGACC AGCAGATCCT GACCGCGCTG CCGCAGCCG AGCGGGCGCT
59281 CATCCCGCTC TACACCGCGT TCGCCCGTGA GGAGTGAGCC TCTCTCGGC GCAGCCGAGC
59341 GGCGGCGTGC CGGTTGTCC CTCTTCGAA CCATGACCG AGCCCGCGCC GGTCCGCGCA
50 59401 GCGGCTAGCG CCGCTCGAG CAGAGAGCG TGGAGCGACA GGCGACGACC CCGCCGAGCG
59461 TGTGAAACGG ATTGCGGAG CCTCATTCG GGATCCCTC CAGACACTCG TTCAGCGCCT
59521 TGGCGTCGAT GCGGCTGGG CACTCGCCGA TGGTCAGCTC GTCGCGCCAG TCGGATCGGA
45 59581 TCTTGTTCGA GCACGCATCC TTGCTCGAAT ACTCCCGTC TTGTCCGATG TTGTTGCACC
59641 GCGCCTCGCG GTGCAACCG GCCGCCACGA TGCTATCGAC GGCGCTGCCG ACTGGCACCG
55 59701 GCGCCTCGCC TTGCGCGCA CCGGGGTTT GCGCCTCCC GCGTACCGC TTTTCGCGC
59761 CCGACGCCG CCGGAGCAG CTCATTCCC ACATCGAGAT CAGGCCACG ACCAGTTTC
59821 CAGCAATCTT TTGCATGGCT TCCCTCCCT CACGACAGT CACATCAGAG ATTCTCCGCT
59881 CCGCTCGTC GTTCGACAG CCGGACGCG CACGAGAGA ACCGTCCCC ACCAGAACAG
50 59941 CCGCATCGG GTTCTCGCA GCATGCCAG ACATCCTTG GACTAGCGT CCTCCGCTG

5
60001 TGCCGAGATC GGCTGTCTCTG TGCGACGGCA ATGTCTCTGCG ATCGGCCGGG CAGGATCGAC
60061 CGACACGGGC GCCGGGCTGG AGGTGCCGCC ACGGGCTCGA AATGCGCTGT GGCAGGCGCC
60121 TCCATGCCCG CTGCCGGGAA CGCAGCGCCC GGCCAGCCTC GGGGCCACGC TCGAACGGG
5 60181 AGATGCTCCC GGAGAGGCGC CGGGCACAGC CGAGCGCCGT CACCACCGTG CGCACTCGTG
60241 AGCGCTAGCT CCTCGGCATA GAAGAGACCG TCACTCCCGG TCCGTGTAGG CGATCGTGCT
60301 GATCAGCGCG TCCTCCGCCT GACGCGACTC GAGCCGGGTA TGCTGCACGA CGATGGGCAC
10 60361 GTCCGATTTC ATCAGCTTGG CATAGTCCGT ATCGCGCGGG ATCGGCTCGG GGTCCGTCAG
60421 ATCGTGAAC CGGACGTGCC GGGTGGCCCT CGCTGGAACG GTCACCCCGT ACCGCCCGGC
60481 GGGTTCGCGG TCGCTGAAGT AGACGGTGAT GGGACCTGC GCGTCCCGGT CCGACGCATT
10 60541 CAACAGGCAG GCCGTCTCAT GGCTCGTCAT CTGCGGCTCA GGTCCGTTGC TCCGGCCTGG
60601 GATGTAGCCC TCTGCGATTG CCCAGCGCCT CCGCCCGATC GGCTTGTTCA TGTGTCTCTC
60661 CTCCTGGCTC CTCTTTGGCA GCCTCCCTCT GCTGTCCAGG TCGACGGCC TCTTCGCTCG
15 60721 ACGCGTTCGG GGCTCCATGG CTGAGAATCC TCGCCGAGCG CTCCTTSCCG ACCGGCGCGC
60781 TGAGCGCCGA CGGGCCTTGA AAGCACGCGA CCGACACGG GATGCCGCG CGACGAGGCC
15 60841 GCCCGCGCTC TGATCCCGAT CGTGGCATCA CGACGTCCGC CGACGCTCG GCAGGCCGGC
60901 GTGAGCGCTG CGCGGTCATC GTCTCTCTCG CGTCACCGCC ACCCGCCGAT TCACATCCCA
60961 CCGCGGCACG ACGCTTGCTC AAACCGCGAC GACACGGCCG GCGGCTGTG GTACCGGCCA
61021 GCCCGGACGC GAGCCCCGAG AGGGACAGTG GGTCCGCGT GAAGCAGAGA GGCATCGAG
20 61081 GTGGTGAGAT GAAACACGTT GACACGGGCC GACGAGTCGG CCGCCGATA GGGCTCACGC
61141 TCGGTCTCCT CGCGAGCATG GCGCTCGCGG GCTGCGGCG CCGAGCGAG AACACCGTGC
20 61201 AGGGCACGCG GCTCGCGCCC GGCCCGATG CGCACGTCAC CGCCGACGTC GACGCCGACG
61261 CCGCGACCAC GCGGCTGGCG GTGGACGTG TTACCTCTC GCGCCCGAG CGGATCGAGG
61321 CCGCGAGCGA GCGGTTCTGTC GTCTGGCAGC GTCCGAATC CCACTCCCG TGGCTACGGG
61381 TCGAGTGCT CCACTACAAC GCTGCCAGCC GAAGAGGCAA GCTGGCCGAG ACGACCGTGC
25 61441 CGCATGCCAA CTTGAGCTG CTCATACCG TCGAGAAGCA GAGCAGCCCT CAGTCGCCAT
61501 CGTCTGCCG CGTCATCGGG CCGACCTCCG TCGGTAACA TCGCGCTATC AGCAGCGCTG
25 61561 AGCCCGCCAG CATGCCCCAG AGCCTGCTT CGATCGCTT CCGCATCATC CGTGGCGACT
61621 CCTCCAGCGA CGGCCGCTC AAAACAACCG CCGTCCCGG GCGGCTCTAC GTGCGGACA
61681 GGAGAGCGTC CTAGCGCGGC CTGCGCATCG CTGGAAGGAT CCGCGGAGCA TGGAGAAAGA
30 61741 ATCGAGGATC GCGATCTACG GCGCCGTGCG CGCCAACGTG GCGATCGCG CCGTCAAGTT
61801 CATCGCCGCC GCCGTGACCG GCAGCTCTGC GATGCTCTCC GAGGCGTGC ACTCCCTCGT
61861 CGATACCGCA GACGGGCTCC TCCTCTGCT CGGCAAGCAC CGGAGCGCCC GCGGCGCGA
30 61921 CGCCGAGCAT CCGTTCGSCC ACGGCAAGGA GCTCTATTTC TGGACGCTGA TCGTCGCCAT
61981 CATGATCTTC GCGCGGSCG GCGGCGTCTC GATCTACGAA GGGATCTTGC ACCTCTTGCA
35 62041 CCGCGCTCG ATCGAGGATC CGACGTGGA CTACGTTGTC CTGCGCGAG CCGCGCTCTT
62101 CGAGGGGACG TCGCTCGCCA TCTCGATCCA CGAGTTCAAG AAGAAAGACG GACAGGGCTA
62161 CGTCCGCGG ATGCGGTCCA GCAAGGACCC GACGACGTT ACGATCGTCC TGGAGGATTC
62221 CGCGGCGCTC GCGGCGCTCG CCATCGCCTT CCTCGGCGTC TGGCTTGGG ACCGCTGGG
35 62281 AAACCCCTAC CTCGACGGCG CGGCGTCGAT CCGCATCGGC CTCGTGCTCG CCGCGGTGCG
40 62341 GGTCTTCTCT GCCAGCCAGA GCCGTGGACT CCTCGTAGGG GAGAGCGCGG ACAGGGAGCT
62401 CCTCGCCGCG ATCCGCGCGC TCGCCAGCGC AGATCCTGGC GTGTGCGCG TGGGGCGGCC
62461 CCTGACGATG CACTTCGGTC CGCACGAAGT CCTGGTCTG CTGCGCATCG AGTTCGACGC
62521 CGCGCTCACG GCGTCCGGGG TCGCGGAGGC GATCGAGCGA ATCGAGACAC GGATACGGAG
45 62581 CGAGCGACCC GACGTGAACC ACATCTACGT CGAGGCCAGG TCGCTCCACC AGCGCGCGAG
62641 GCGGTGACGC GCCGTGGAGA GACCGCTCGC GGCTCCGCG ATCTCCGCG GCGCCGGGC
40 62701 TCGGTTAGCC CTCGAGCAG GCGCGCGCTG GCGGCAAAAC CGTGAAGACG TCGTCTCTG
62761 ACGCGAGGTA CGCTGGTTGC AAGTTGTAC GCCGTATCG GAGGTCCGGC AGCGCCGGAG
62821 CCGGGCGCGT CCGGCGGCAC GAAGGCCCG CGAGCGCGG CTTGAGGGG GCGACGTCAT
50 62881 GAGGAAGGGC AGGGCGCATG GGGCGATGCT CCGCGGGCGA GAGGACGGCT GCGCTCGCG
62941 CCTCCCGGCG GCCGCGCGC TTCGCGCCGC GCTCCAGCG GGTGCTCGC GCGATCTCG
63001 CCGCGCGCGG CTCATCGCG CCGTGTCCCT CACCGCGGC GCCAGCATG CCGTCTCTC
45 63061 GCTGTTCAG CTCGGATCA TCGAGCACCT GCGGATCCT CCGCTTCCAG GGTTCGATT
63121 GGCCAAGGTG ACGAGCTCCG ATATCGCGTT CCGGCTCACG ATGCCGAGC GCGCGCTCG
63181 GCTCACCAGC TTCGCTCCA ACCTGGCGCT GGCTGGCTGG GGAGGCGCG AGCGCGCCAG
55 63241 GAACACCCCG TGGATCCCG TCGCGTGGC GGCCAAGGCG GCCGTGAGG CCGCGGTGTC
63301 CGGATGGCTC CTCGTCCAGA TCGACGGCG GGAGAGGGCC TGGTSCGCGT ACTGCCTGCT
63361 CGCCATGGCG GCAACATGG CCGTGTTCGC GCTCTGCTC CCGGAAGGGT GGGCGGCGCT
63421 GAGGAAGGCG CGAGCGCGCT CGTGACAGG CCGTGGCGG GCGCGGCCA TCGGAGGCCG
50 63481 CCGTGACCC GCTCCGTAC GCCCGGCCC GCGCGCGGT GAGCTGCCGC GGACAGGGCG

5 63541 CGTACCGTGG ACCCCGACAG CGCCGCGTCG ACGGACATCC CCGGCGGCTC GCGCGGCGCG
63601 GCCGCGCAA CTCCGGCCCG CCGCCGGGCA TCGACATCTC CCGCGAGCAA GGGUACTCCG
63661 CTCCTGCCCC CGTCCGCGAA CGATGGCTGC GCTGTTTCCA CCCTGGAGCA ACTCCGTTTA
63721 CCGCGTGGCG CTCGTCGGGC TCATCGCCTC GCGGGGCGGC GCCATCCTCG CGCTCATGAT
5 63781 CTACGTCCGC ACGCCGTGGA AGCGATACCA GTTCGAGCCC GTCGATCAGC CCGTGCAGTT
63841 CGATACCGGC CATCACGTGC AGCAGGATGG CATCGATTGC GTCTACTGCC ACACCACGGT
10 63901 GACCCGCTCG CCGACGGCGG GGATGCCGCC GACGGCCACG TGCATGGGGT GCCACAGCCA
63961 GATCTGGAAT CAGAGCGTCA TGCTCGAGCC CGTGGCGCGG AGCTGGTTCT CCGGCATGCC
10 64021 GATCCCGTGG AACCGGGTGA ACTCCGTGCC CGACTTCGTT TATTTCAACC ACGCGATTCA
64081 CGTGAACAAG GCGGTGGGCT GCGTGAGCTG CCACGGGCGC GTGGACGAGA TGGCGGCCGT
64141 CTACAAGGTG GCGCCGATGA CGATGCCCTG GTGCGTGGAG TGCCATCGCC TGCCGGAGCC
64201 GCACCTGCGC CCGCTCTCCG CGATACCGA CATGCGCTGG GACCCGGGGG AACGGAGGGA
15 64261 CGAGCTCGGG GCGAAGCTCG CGAAGGAGTA CCGGCTCCCG CCGCTCACGC ACTGCACAGC
64321 GTGCCATCGA TGAACGATGA ACAGGGGATG TCCGTGAAAG ACGCAGATGA GATGAAGGAA
15 64381 TGGTGGCTAG AAGCGCTCGG GCGGCGGCGA GAGCGCGCGT CCTACAGGCT GCTGGCGCCG
64441 CTCATCGAGA CCCCCGAGCT CCGCGCGCTC CCGCGGCGG AACC GCCCGG GGGCGTGGAC
64501 GAGCCGGCGG GCGTCAGCCG CCGCGCGCTG CTCAGCTGC TCGGCGCGAG CATGGCGCTC
64561 GCGGCGCTCG CCGGCTGCAC CCGCGCATGAG CCGGAGAAGA TCCTGCCGTA CAACGAGACC
20 64621 CCGCCCGGCG TCGTGCCGGG TCTCTCCAG TCCTACGCGA CGA3CATGGT GCTCGACGGG
20 64681 TATGTCATGG GCCTCCTCGC CAAGAGCTAC GCGGGGCGGC CCATCAAGAT CGAGCCCAAC
64741 CCGCGCGACC CCGCGAGCCT CCGCGCGACC GCGCTCCAG AGCAGGCCTC GATCCTCTCG
64801 CTGTACGACC CGTACCGCGC GCGCGCGCCG ACGCGCGCGG GCCAGGTCGC GTCGTGGGAG
64861 GCGCTCTTGG CCGGCTTCGG CCGCGACCGC GAGGACCCCG ECGCTGGCCT CCGCTTCGTC
25 64921 CTCCAGCCCA CGAGCTCGCC CCTCATCGCC SCGCTGATCG AGCGCGTCCG GCGCAGGTTT
64981 CCGGCGCGCG GGTTCACCTT CTGGTCGCCG GTCCACGCCG AGCAAGCGCT CGAAGGCGCG
25 65041 CCGGCGCGCG TCGGCTTCAG GCTCTTCCT CAGCTCGACT TCGACCAGGC CGAGGTGATC
65101 CTCGCCCTGG ACGCGGACTT CCTCGCGGAC ATGCCGTTCA GCGTGCGCTA TCGCGCGGAC
65161 TTCGCCGCGC GCGGCGGACC CCGGAGCCCG GCGGCGGCCA TGAACCGCCT CTACGTCCCG
65221 GAGGCGATGT TCACGCCCAC GGGGACGCTC GCGGACCACC GGCTCCGCGT SCGCGCCGCG
30 65281 GAGGTGCGCG GCGTCGCGG CCGGCTCGCG GCGGAGCTCG TGCACGGCCT CCGCCTGCGC
65341 CCGCGCGGGA TCACGGACGC CGACGCCGCC GCGCTGCGCG CGCTCCGCCC CCGGACGGC
65401 GAGGGGACAG GCGCCTTCGT CCGGCGCTC GCGCGCGATC TCGCGCGCGC GGGGGGCGCC
30 65461 GCGGTCGCGG TCGTGGCGA CCGGCGAGCC CCCATCGTCC ACGCCCTCCG GCACGTCATC
65521 AACGCCGCGC TCCGACGCCG GCGGCGCTGG ATGGTCGATC CTGTGCTGAT CGACGCGGGC
35 65581 CCCTCCACGC AGGGCTTCTC CGAGCTCGTC GCGGAGCTCG GGCGCGGCGC GGTGACACCC
65641 TGATCCTCCT CGAGCTGAAC CCGGTGTACG CCGGCGCGGC CGACGTCGAT TTCGCGGGCC
65701 TCCTCGCGCG C3TGCCACG AGCTTGAAG CCGG3CTCTA CGACGACGAG ACCGCCCGCG
65761 CTTGCACGTG GTTCGTGCGG ACCCGGCATT ACCTCGAGTC GTGGGGGAC GCGCGGCGCT
35 65821 ACGACGGGAC GGTCTCGTTC GTGCAACCCC TCGTCCGGCC GCTGTTGAC GGCCGGGCGG
40 65881 TGCCCGAGCT GTCGCGGTC TTCGCGGGG ACGAGCGCCC GGATCCCGG CTGCTGCTGC
65941 GCGAGCACTG GCGCGGCGCG CCGGAGAGG CCGATTTCGA GGCTTCTGG GCGGAGGCAT
66001 TGAAGCGCGG CTTCCTCCCT GACAGCGCCC GCGCGAGGCA GACACCGGAT CTCGCGCCCG
66061 CCGACCTCGC CAAGGAGCTC GCGCGGCTCG CCGCGCGCGC GCGCGCGGCC GGCCCGCGCG
45 66121 TCGACGTGCC GTTCCTCAGG TCGCGCTCGG TCCACGACGG CAGGTTCCGC AACAAACCCCT
66181 GGCTGCAAGA GTCGCCGCG CCGATACCA GGCTCACCTG GGGCAACGCC GCCATGATGA
40 66241 GCGCGGCGAC CCGGCGCGG CTCGGCGTCG AGCCCGGCGA TGTCTCGAG CTCGCGCTGC
66301 GCGGCCGTAC GATCGAGATC CCGGCCGTCG TCGTCCGCGG GCACGCGGAC GACGTGATCA
66361 GCGTCGACCT CCGCTACGGG CCGGAGGAGT CCGGAGGAGT CCGGCGCGG GTGGCGGTST
66421 CCGCGTATCG GATCCGCCCG TCCGACGCGC GGTGGTTGCG GGGGGGCGTC TCCGTGAGGA
50 66481 AGACCGGCGC CACGGCCGCG CTCGCGCTGG CTCAGATCGA GCTGTCCAG CACGACCGTC
66541 CCATCGCGCT CCGGAGGAGC CTGCGCGAGT ACCGTGAACA GCGCGGTTTC GCGGAGGAGC
45 66601 ACAAGGGGCG GGTCCGCTCG ATCCTGCCGG AGGTGAGTA CACCGGCGCG CAATGGGCGA
66661 TGTCATCGA CATGTGATC TGCACCGGGT GCTCCTCGTC CCTCGTGGCC TGTCAGGCCG
66721 AGAACAACGT CCTCGTCGTC GGCAAGGAGG AGGTGATGCA CCGCGCGAG ATGCAGTGGT
55 66781 TGCGGATCGA TCAGTACTTC GAGGGTGGAG GCGACGAGGT GAGCGTCGTC AACCAGCCGA
66841 TGCTCTGCCA GCACTGCGAG AAGGCGCCGT GCGAGTACGT CTGTCCGGTG AACGCGACGG
66901 TCCACAGCCC CGATGGCCTC AACGAGATGA TCTACAACCG ATGCATCGGG ACGCGCTTTT
66961 GCTCCAACAA CTGTCCGTAC AAGATCCGGC GGTTCATTT CTTCGACTAC AATGCCACG
50 67021 TCCCGTACAA CCGCGGCTC CCGAGGCTCC AGCGCAACCC GGACGTCACC GTCCGCGCCC

5 67081 GCGGCGTCAT GGAGAAATGC ACGTACTGCG TGCAGCGGAT CCGAGAGGCG GACATCCGCG
67141 CGCAGATCGA GCGGCGGCGG CTCGGGCGCG GCGAGGTGGT CACCGCCTGC CAGCAGGCCT
67201 GTCCGACCGG CGCGATCCAG TTCGGGTGCG TGGATCAUGC GGATACAAAG ATGGTCGCGT
5 67261 GCGGAGGGA GCGGCGCGCG TACGCGTGC TCCACGACCT CGGCACCCGG CCGCGGACGG
67321 AGTACCTCGC CAAGATCGAG AACCCGAACC CGGGGCTCGG GCGGAGGGG GCGGAGAGGC
67381 GACCCGGAGC CCGGAGCGTC AAACCCGCGC TCGGGGCGGA CGGCGCCGAG AGGCGACCCG
10 67441 GAGCCCCGAG CGTCAAACCC GAGATTGAAT GAGCCATGGC GGGCCCGCTC ATCCTGGAUG
67501 CACCGACCGA CGATCAGCTG TCGAAGCAGC TCCTCGAGCC GGTATGGAAG CCGCGCTCCC
10 67561 GGCTCGGCTG GATGCTCGCG TTCGGGCTCG CGCTCGGCGG CACGCGCTG CTCTTCTCG
67621 CGATCACCTA CACCGTCCTC ACCGGGATCG GCGTGTGGG CAACAACATC CCGGTGCGCT
67681 GGGCCTTCGC GATCACCAAC TTCGTCTGGT GGATCGGGAT CGGCCACGCC GGGACGTTCA
67741 TCTCCGCGAT CCTCCTCCTG CTCGAGCAGA AGTGCGGAC GAGCATCAAC CGCTTCGCCG
15 67801 AGCGATGAC GCTCTTCGCG GTCGTCCAGG CCGGCTCTT TCCGGTCTC CACCTCGGCC
67861 GCGCCTGGTT CGCTACTGAG ATCTTCCCGT ACCCGCGGAC GATGCAGGTG TGGCTCGAGT
15 67921 TCCGGAGCGC GCTGCGTGG GACGCGCGCG CGATCGCGAC CTACTTCAGG GTGTGCTCC
67981 TGTTCTGGTA CATGGGCTC GTCCCGGATC TGGCGCGCT GCGCGACCAC GCGCGGGCG
68041 GCGTCCGGCG GGTGATCTAC GGGCTCATGT CGTTCGGCTG GCACGGCGCG GCGGACCACT
68101 TCCGCGATTA CCGGGTGCTG TACGGGCTGC TCGCGGGGCT CGCGACGCC CTGCTCGTCT
20 68161 CGGTGCACTC GATCGTGAGC AGCGATTTCG CGATCGCCCT GGIGCCCGGC TGGCACTCGA
68221 CGCTCTTTC GCGGTCTTTC GTGCGGGCG CGATCTTCTC CGGGTTCGCG ATGGTCTCA
20 68281 CGCTGCTCAT CCGGTGCGG CGGATCTACG GGCTCCATAA CGTCGTGACC GCGCGCCACC
68341 TCGACGATCT CGCGAAGATG ACGCTCGTGA CCGGCTGGAT CGTCATCTC TCGTACATCA
68401 TCGAGAACTT CCTCGCCTGG TACAGCGGCT CCGGCTACGA GATGCATCAG TTTTTCAGA
25 68461 CGCGCCTGCA CCGCCCCAAC AGCGCGGCT ACTGGGCGCA GCACGTCTG AACGTGCTCG
68521 TCATCCAGCT CCTCTGGAGC GAGCGGATCC GGACGAGCCC CGTCGCGCTC TGGCTCATCT
68581 CCTCCTGGT CAACGTGCGG ATGTGGAGCG AGCGGTTCAC GCTCATCGTG ATGTGCTCG
25 68641 AGCAAGAGTT CCTCCCGTCC AAGTGGCAGG GCTACAGCCC GACGTGGGTG GACTGGAGCC
68701 TCTTCATCGG GTCAGGCGGC TTCTTCATGC TCCTGTTCTT GAGCTTTTG CGCGTCTTC
68761 CGTTCATCCC CGTTCGCGAG GTCAAGGAGC TCAACCATGA AGAGCTGGAG AAGGCTCGGG
30 68821 GCGAGGGGGG CCGCTGATGG AGACCGGAAT GCTCGGCGAG TTCGATGACC CGGAGGCGAT
68881 GCTCCATGCG ATCCGAGAGC TCAGGCGGCG CCGCTACCGC CGGCTGGAAG CGTTCACGCC
68941 CTATCCGGTG AAGGGGCTCG ACCAGGCGCT CCGCCTCCCG CGCTCGAACC TCAACCGGAT
30 69001 GCTGCTGCCC TTCGCGATCC TGGGGGTCGT GGGCGGCTAC TTCGTCCAGT GGTCTGCAA
69061 CGCTTTCCAC TATCCGCTGA ACGTGGGCGG GCGCCCGCTG AACTCGGGCG CCGGCTTCAT
35 69121 CCGGATCAGC TTCGAGATGG GGGTGTCTC CACCTCGATC TTCGGCGTGC TCATCGGCTT
69181 TTACCTGACG AGGCTGCCGA GGCTCTACCT CCGCTCTTC GACGCCCCGG GCTTCGAGCG
69241 CGTCACGCTG GATCGGTTTC TGGTCCGGCT CGACGACACG GAACCTTCCT TCTCGAGCGC
69301 CCAGGCGGAG CGCGACCTCC TCGCGCTCGG CGCCCGGCGC GTCGTGTCG CGAGGAGGCG
35 69361 CGAGGAGCCA TGAGGGCCGG CGCCCGGCT CCGCCTCTCG GCGCGCGCT CGCGCGGTT
69421 GCGCTCGTCC TGCTCGCCGG GTGCGCGAG AAGGTGCTG CCGAGCCGGA CTTCGAGCGG
69481 ATGATCCGCC AGGAGAAATA CGGACTCTGG GAGCCGTGCG AGCACTTCGA CGACGCGCGC
69541 GCGATGCAGC ACCCGCCCGA GGGGACCGTC GCGCGCGGGC GCGTACCGG GCGCCCGGC
69601 TATCTCCAGG GCGTCTCGA CCGGGCGTAC GTACGGAGG TGCCGCTCTT GCTCACGCTC
45 69661 GAGCTCGTGC AGCGCGGCG GCGAGCGTTC GAGACCTTCT GCGCGCGGTG CCACGGGATC
69721 CTCGGCGACG GCAGCTCGCG CGTGGCGAG AACATGACGC TCGCGCGGCC CCGTCTGCTC
40 69731 ATCGGACCCG AGGCGCGGAG CTTCCCGCG GCGAGGATCT ACCAGTTCAT CATCGAGGGC
69841 TACGGCCTGA TGCCGCGCTA CTCGGACGAT CTGCCCCGCA TCGAAGAGCG CTGGGCGGTG
69901 CTCGCTACG TGAAGGCGCT TCAGCTGAGC CGCGGAGTGG CCGCGGGCGC CCTCCCGCCA
50 69961 GCGCTCCGCG GCGGGGCGA GCAGGAGCTG CGATGAACAG GGATGCCATC GAGTACAAGG
70021 GCGGCGCGAC GATCGCGGCC TCGCTCGCGA TCGCGGCGCT CGGCGCGGTC GCGCGATCG
70081 TCGGCGGCTT CGTCGATCTC CGCCGCTCT TCTTCTCGTA CCTCGCCGCG TGGTCGTTCT
45 70141 CCGTGTCTCT GTCCGTGGGC CCGCTCTGCA CGCTCTCAC CTGCAACGCC ATGCGCGCGG
70201 GCTGGCCAC GCGGCTGCGC CGCTCTCTCG AGACGATGGT GCGCGCGCTG CCTCTGCTCG
70261 CCGCGCTCTC CGCGCCGATC CTGGTCCGCC TGGACACGCT GTATCCGTGG ATGACCCCG
55 70321 AGCGGATCGC CCGCGAGCAC GCGCGGCGCA TCCTCGAGCA CAGGGCGCCC TACTTCAATC
70381 CAGGCTTCTT CGTCGTGCGC TCGGCGATCT ACTTCGCGAT CTGGATCGCC GTCGCGCTCG
70441 TGCTCCGCG GCGATCGTTC GCGCAGGACC GTGAGCCGAG GGCCGACGTC AAGGACGCGA
70501 TGATGGCCT GAGCGGCGCC ATGCTGCGCG TCGTGGCGAT CACGATCGTC TTCTCGTCTG
50 70561 TCGACTGGCT CATGTCCCTC GACGCGACCT GGTACTCGAC GATGTTCCCG GTCTACGTT

5 70621 TCGCGAGCGC CTTCGTGACC GCCGTCGGCG CGCTCAGGGT CCTCTCGTAT GCCGCGCAGA
70681 CGTCCGGCTA CCTCGCGAGG CTGAACGACT CGCACTATTA CGCGCTCGGG CGGCTGCTCC
70741 TCGCGTTCAC GATATTCTGG GCCTATGCGG CCTATTTCCA GTTCATGTTG ATCTGGATCG
5 70801 CGAACAAGCC CGATGAGGTC GCCTTCTTCC TCGACCGCTG GGAAGGGCCC TGGCGGCCGA
70861 CCTCCGTGCT CGTCGTCTTC ACGCGGTTTC TCGTCCCGTT CCGATCCTG ATGTCGTACG
70921 CGATCAAGCG GCGCCCGCGC CAGCTCTCGT GGATGGCGCT CTGGGTCTGC GTCTCCGGCT
10 70981 ACATCGACTT TCACTGGCTC GTGGTGCCGG CGACAGGCG CCACGGGTTC GCCTATCACT
71041 GGCTCGACCT CGCGACCCTG TCGGTCTGGG GCGGCCTCTC GACCGCGTTC GCCGCGTGGC
71101 GGCTCGAGG GCGGCCGGTG GTCCCGTCC ACGACCGCG GCTCGAAGAG GCCTTTGCGT
10 71161 ACCGGAGCAT ATGATGTTCC GTTCCGTC CAGCGAGGTT CGCCAGGAGC AGGACACGCT
71221 CCCCTGGGGG CGCGTGATCC TCGCGTTCGC CGTCGTGCTC GCGATCGGCG GCGCGCTGAC
71281 GCTCTGGGCC TCGCTCGCGA TCGGGGCCCC CGAGGCGGAT CTGCGGCCCT CCCTCGCGTT
15 71341 CCCCAGAGAAG GATCTCGGGC CGC3GCGCGA GGTGCGCATG GTCCAGCAGT CGCTGTTTCA
71401 CGAGCGCGCG CTGGGCCAGC AGCTCGTCGA CGCGCAGCGC CCGGAGCTCC GCCGCTTCGG
15 71461 CCTCGTCGAT CGGGAGAGGG GCATCGTGAG CATCCCGATC GACGACGCGA TCGAGCTCAT
71521 GGTG3CGGGG GCGCGCGGAT GAGCCGG3CC GTCGCCGTGG CCCTCCTGCT GGCAGCCGGC
71581 CTCGTGTGCG GCCCGGGCGC CGCGTCCGAG CCCGAGCGCG CGCGCCCCGC GCTGGGCCCG
71641 TCGCGGGCCG ACGCCGCGCC GCGGAGCGAC GGCTCCGGCG CGGAGGAGCC GCCCGAAGGC
20 71701 GCCTTCCTGG AGCCACCGC CGGGGTGGAC ATCGAGGAGC GCCTCGGCCG CCCGGTGGAC
71761 CGCGAGCTCG CCTTCACCGA CATGGACGGG CGGCGGGTGC GCCTCGCGCA CTACTTCGCC
20 71821 GACGGCAAGC CCTCCTCCT CGTCCTCGCG TACTACCGGT GTCCCGCGCT GTGCGGCCTC
71881 GTGCTGCGCG GCGCCGTGCA GGGGCTGAAG CTCCTCCCGT ACCG3CTCGG CGAGCAGTTC
71941 CACGCGCTCA CGGTCAGCTT CGACCCGCGC GAGCGCCCCG CGGCCGDD

Example 2

Construction of a *Myxococcus xanthus* Expression Vector

The DNA providing the integration and attachment function of phage Mx8 was inserted into commercially available pACYC184 (New England Biolabs). An ~2360 bp MfeI-SmaI from plasmid pPLH343, described in Salmi *et al.*, Feb. 1998, J. Bact. 180(3): 614-621, was isolated and ligated to the large EcoRI-XmnI restriction fragment of plasmid pACYC184. The circular DNA thus formed was ~6 kb in size and called plasmid pKOS35-77.

Plasmid pKOS35-77 serves as a convenient plasmid for expressing recombinant PKS genes of the invention under the control of the epothilone PKS gene promoter. In one illustrative embodiment, the entire epothilone PKS gene with its homologous promoter is inserted in one or more fragments into the plasmid to yield an expression vector of the invention.

The present invention also provides expression vectors in which the recombinant PKS genes of the invention are under the control of a *Myxococcus xanthus* promoter. To construct an illustrative vector, the promoter of the *pilA* gene of *M. xanthus* was isolated as a PCR amplification product. Plasmid pSWU357, which comprises the *pilA* gene promoter and is described in Wu and Kaiser, Dec. 1997, J. Bact. 179(24):7748-7758, was mixed with PCR primers Seq1 and Mxpil1 primers:

Seq1: 5'-AGCGGATAACAATTCACACAGGAAACAGC-3'; and
Mxpil1: 5'-TTAATTAAGAGAAGGTTGCAACGGGGGGC-3',
and amplified using standard PCR conditions to yield an ~800 bp fragment. This fragment
was cleaved with restriction enzyme KpnI and ligated to the large KpnI-EcoRV restriction
fragment of commercially available plasmid pLitmus 28 (New England Biolabs). The
resulting circular DNA was designated plasmid pKOS35-71B.

The promoter of the *pilA* gene from plasmid pKOS35-71B was isolated as an ~800
bp EcoRV-SnaBI restriction fragment and ligated with the large MscI restriction fragment
of plasmid pKOS35-77 to yield a circular DNA ~6.8 kb in size. Because the ~800 bp
fragment could be inserted in either one of two orientations, the ligation produced two
plasmids of the same size, which were designated as plasmids pKOS35-82.1 and pKOS35-
82.2. Restriction site and function maps of these plasmids are presented in Figure 3.

Plasmids pKOS35-82.1 and pKOS35-82.2 serve as convenient starting materials
for the vectors of the invention in which a recombinant PKS gene is placed under the
control of the *Myxococcus xanthus pilA* gene promoter. These plasmids comprise a single
PacI restriction enzyme recognition sequence placed immediately downstream of the
transcription start site of the promoter. In one illustrative embodiment, the entire
epothilone PKS gene without its homologous promoter is inserted in one or more
fragments into the plasmids at the PacI site to yield expression vectors of the invention.

The sequence of the *pilA* promoter in these plasmids is shown below.

```
CGACGCAGGTGAAGCTCCTTCGTGTGCTCCAGGAGCGGAAGGTGAAGCCGGTCCGGCAGCGCCGCGGAGATTCC
CCTTCCAGGCGCGTGTTCATCGCGGCAACGAACCGGCGGCTCGAAGCCGAAGTAAAGGCCGACGCTTTCGTG
AGGACCTCTTCTACCGGCTCAACGTCATCAGTTGGAGCTGCCTCCACTGCGCGAGCGTTCCGGCGACGTTG
CGTTGCTGCGCAACTACTTCTGTCCAGACTGTTCGGAGGAGTTGGGGCGACCCGGTCTGCGTTTCTCCCCCG
AGACACTGGGGCTATTGGAGCGCTATCCCTTCCAGGCAACGTGCGGCAGCTGCAGAACATGGTGGAGCGGG
CCGCGACCCCTGTTCGGATTACAGCCTCCTGGGGCCCTCCACGCTTCCACCCGAGTGCAGGGCGGATACAGACC
CCGCCGTGCGTCCCGTGGAGGGCAGTGAGCCAGGGCTGGTGGCGGGCTTCAACCTGGAGCGGCATCTCGACG
ACAGCGAGCGGCGCTATCTCGTCCGCGCCATGAAGCAGGCGGGGGCGTGAAGACCCGTGCTGCGGAGTTGC
TGGGCCTTTCGTTCCGTTTCCGTTACCGCTACCGGTTGGCCAGCATGGGCTGACGGATGACTTGGAGCCCGGGA
GCGCTTCGGATGCGTAGGCTGATCGACAGTTATCGTCAGCGTCACTGCCGAATTTTGTGAGCCCTGGACCCA
TCCTCGCCGAGGGGATTGTTCCAAGCCTTGAGAATTGGGGGGCTTGGAGTGCGCACCTGGGTGGCATGCGT
AGTGCTAATCCCATCCGCGGGCGCAGTGCCCCCGTTGCAACCTTCTCTTAATTAA
```

To make the recombinant *Myxococcus xanthus* host cells of the invention,
M. xanthus cells are grown in CYE media (Campos and Zusman, 1975, Regulation of
development in *Myxococcus xanthus*: effect of 3': 5'-cyclic AMP, ADP, and nutrition,
Proc. Natl. Acad. Sci. USA 72: 518-522) to a Klett of 100 at 30°C at 300 rpm. The
remainder of the protocol is conducted at 25°C unless otherwise indicated. The cells are
then pelleted by centrifugation (8000 rpm for 10 min. in an SS34 or SA600 rotor) and

5 resuspended in deionized water. The cells are again pelleted and resuspended in 1/100th of the original volume.

10 DNA (one to two μ L) is electroporated into the cells in a 0.1 cm cuvette at room temperature at 400 ohm, 25 μ FD, 0.65 V with a time constant in the range of 8.8 - 9.4. The
5 DNA should be free of salts and so should be resuspended in distilled and deionized water or dialyzed on a 0.025 μ m Type VS membrane (Millipore). For low efficiency
15 electroporations, spot dialyze the DNA, and allow outgrowth in CYE. Immediately after electroporation, add 1 mL of CYE, and pool the cells in the cuvette with an additional 1.5
10 mL of CYE previously added to a 50 mL Erlenmeyer flask (total volume 2.5 ml). Allow the cells to grow for four to eight hours (or overnight) at 30 to 32°C at 300 rpm to allow
20 for expression of the selectable marker. Then, plate the cells in CYE soft agar on plates with selection. If kanamycin is the selectable marker, then typical yields are 10^3 to 10^5 per
25 μ g of DNA. If streptomycin is the selectable marker, then it must be included in the top agar, because it binds agar.

15 With this procedure, the recombinant DNA expression vectors of the invention are electroporated into *Myxococcus* host cells that express recombinant PKSs of the invention and produce the epothilone, epothilone derivatives, and other novel polyketides encoded thereby.

20 Example 3

Construction of a Bacterial Artificial Chromosome (BAC) for Expression of Epothilone in 35 *Myxococcus xanthus*

To express the epothilone PKS and modification enzyme genes in a heterologous
40 host to produce epothilones by fermentation, *Myxococcus xanthus*, which is closely related to *Sorangium cellulosum* and for which a number of cloning vectors are available, can also
25 be employed in accordance with the methods of the invention. Because both *M. xanthus* and *S. cellulosum* are myxobacteria, it is expected that they share common elements of
45 gene expression, translational control, and post translational modification (if any), thereby enhancing the likelihood that the epo genes from *S. cellulosum* can be expressed to
30 produce epothilone in *M. xanthus*. Secondly, *M. xanthus* has been developed for gene cloning and expression. DNA can be introduced by electroporation, and a number of
50 vectors and genetic markers are available for the introduction of foreign DNA, including those that permit its stable insertion into the chromosome. Finally, *M. xanthus* can be

5 grown with relative ease in complex media in fermentors and can be subjected to manipulations to increase gene expression, if required.

To introduce the epothilone gene cluster into *Myxococcus xanthus*, one can build the epothilone cluster into the chromosome by using cosmids of the invention and homologous recombination to assemble the complete gene cluster. Alternatively, the complete epothilone gene cluster can be cloned on a bacterial artificial chromosome (BAC) and then moved into *M. xanthus* for integration into the chromosome.

To assemble the gene cluster from cosmids pKOS35-70.1A2, and pKOS35-79.85, small regions of homology from these cosmids have to be introduced into *Myxococcus xanthus* to provide recombination sites for larger pieces of the gene cluster. As shown in Figure 4, plasmids pKOS35-154 and pKOS90-22 are created to introduce these recombination sites. The strategy for assembling the epothilone gene cluster in the *M. xanthus* chromosome is shown in Figure 5. Initially, a neutral site in the bacterial chromosome is chosen that does not disrupt any genes or transcriptional units. One such region is downstream of the devS gene, which has been shown not to affect the growth or development of *M. xanthus*. The first plasmid, pKOS35-154, is linearized with DraI and electroporated into *M. xanthus*. This plasmid contains two regions of the dev locus flanking two fragments of the epothilone gene cluster. Inserted in between the epo gene regions are the kanamycin resistance marker and the galK gene. Kanamycin resistance arises in colonies if the DNA recombines into the dev region by a double recombination using the dev sequence as regions of homology. This strain, K35-159, contains small regions of the epothilone gene cluster that will allow for recombination of pKOS35-79.85. Because the resistance markers on pKOS35-79.85 are the same as that for K35-159, a tetracycline transposon was transposed into the cosmid, and cosmids that contain the transposon inserted into the kanamycin marker were selected. This cosmid, pKOS90-23, was electroporated into K35-159, and oxytetracycline resistant colonies were selected to create strain K35-174. To remove the unwanted regions from the cosmid and leave only the epothilone genes, cells were plated on CYE plates containing 1% galactose. The presence of the galK gene makes the cells sensitive to 1% galactose. Galactose resistant colonies of K35-174 represent cells that have lost the galK marker by recombination or by a mutation in the galK gene. If the recombination event occurs, then the galactose resistant strain is sensitive to kanamycin and oxytetracycline. Strains sensitive to both antibiotics are verified by Southern blot analysis. The correct strain is identified and designated K35-

5 175 and contains the epothilone gene cluster from module 7 through two open reading frames past the *epoL* gene.

10 To introduce modules 1 through module 7, the above process is repeated once more. The plasmid pKOS90-22 is linearized with *DraI* and electroporated into K35-175 to
5 create K35-180. This strain is electroporated with the tetracycline resistant version of pKOS35-70.1A2, pKOS90-38, and colonies resistant to oxytetracycline are selected. This creates strain K35-185. Recombinants that now have the whole epothilone gene cluster are
15 selected by resistance to 1% galactose. This results in strain K35-188. This strain contains all the epothilone genes as well as all potential promoters. This strain is fermented and
10 tested for the production of epothilones A and B.

20 To clone the whole gene cluster as one fragment, a bacterial artificial chromosome (BAC) library is constructed. First, SMP44 cells are embedded in agarose and lysed according to the BIO-RAD genomic DNA plug kit. DNA plugs are partially digested with
restriction enzyme, such as *Sau3AI* or *HindIII*, and electrophoresed on a FIGE or CHEF
25 gel. DNA fragments are isolated by electroeluting the DNA from the agarose or using gelase to degrade the agarose. The method of choice to isolate the fragments is
15 electroelution, as described in Strong *et al.*, 1997, Nucleic Acids Res. 19: 3959-3961, incorporated herein by reference. The DNA is ligated into the BAC (pBeloBACII) cleaved
30 with the appropriate enzyme. A map of pBeloBACII is shown below.

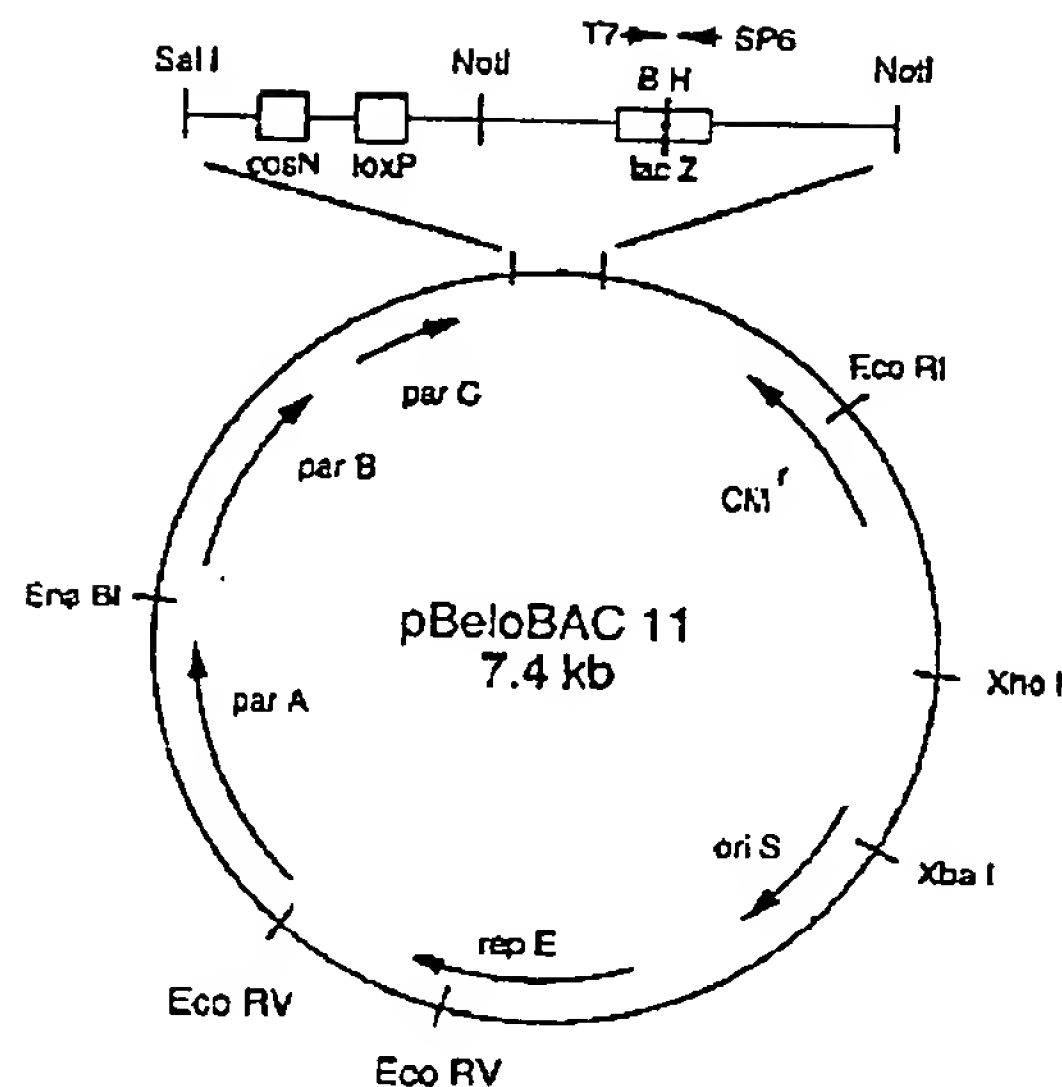
35

40

45

50

55



The DNA is electroporated into DH10B cells by the method of Sheng *et al.*, 1995, Nucleic Acids Res. 23: 1990-1996, incorporated herein by reference, to create an *S. cellulosum* genomic library. Colonies are screened using a probe from the NRPS region of the epothilone cluster. Positive clones are picked and DNA is isolated for restriction analysis to confirm the presence of the complete gene cluster. This positive clone is designated pKOS35-178.

To create a strain that can be used to introduce pKOS35-178, a plasmid, pKOS35-164, is constructed that contains regions of homology that are upstream and downstream of the epothilone gene cluster flanked by the dev locus and containing the kanamycin resistance galK cassette, analogous to plasmids pKOS90-22 and pKOS35-154. This plasmid is linearized with DraI and electroporated into *M. xanthus*, in accordance with the method of Kafeshi *et al.*, 1995, Mol. Microbiol. 15: 483-494, to create K35-183. The plasmid pKOS35-178 can be introduced into K35-183 by electroporation or by transduction with bacteriophage P1 and chloramphenicol resistant colonies are selected. Alternatively, a version of pKOS35-178 that contains the origin of conjugative transfer from pRP4 can be constructed for transfer of DNA from *E. coli* to K35-183. This plasmid

5 is made by first constructing a transposon containing the oriT region from RP4 and the tetracycline resistance maker from pACYC184 and then transposing the transposon *in vitro* or *in vivo* onto pKOS35-178. This plasmid is transformed into S17-1 and conjugated
10 into *M. xanthus*. This strain, K35-190, is grown in the presence of 1% galactose to select
5 for the second recombination event. This strain contains all the epothilone genes as well as all potential promoters. This strain will be fermented and tested for the production of epothilones A and B.

15 Besides integrating pKOS35-178 into the dev locus, it can also be integrated into a phage attachment site using integration functions from myxophages Mx8 or Mx9. A
10 transposon is constructed that contains the integration genes and att site from either Mx8 or Mx9 along with the tetracycline gene from pACYC184. Alternative versions of this
20 transposon may have only the attachment site. In this version, the integration genes are then supplied in trans by coelectroporation of a plasmid containing the integrase gene or having the integrase protein expressed in the electroporated strain from any constitutive
25 15 promoter, such as the mgl promoter (see Magrini *et al.*, Jul. 1999, J. Bact. 181(13): 4062-4070, incorporated herein by reference). Once the transposon is constructed, it is transposed onto pKOS35-178 to create pKOS35-191. This plasmid is introduced into
30 *Myxococcus xanthus* as described above. This strain contains all the epothilone genes as well as all potential promoters. This strain is fermented and tested for the production of
20 epothilones A and B.

35 Once the epothilone genes have been established in a strain of *Myxococcus xanthus*, manipulation of any part of the gene cluster, such as changing promoters or swapping modules, can be performed using the kanamycin resistance and galK cassette.

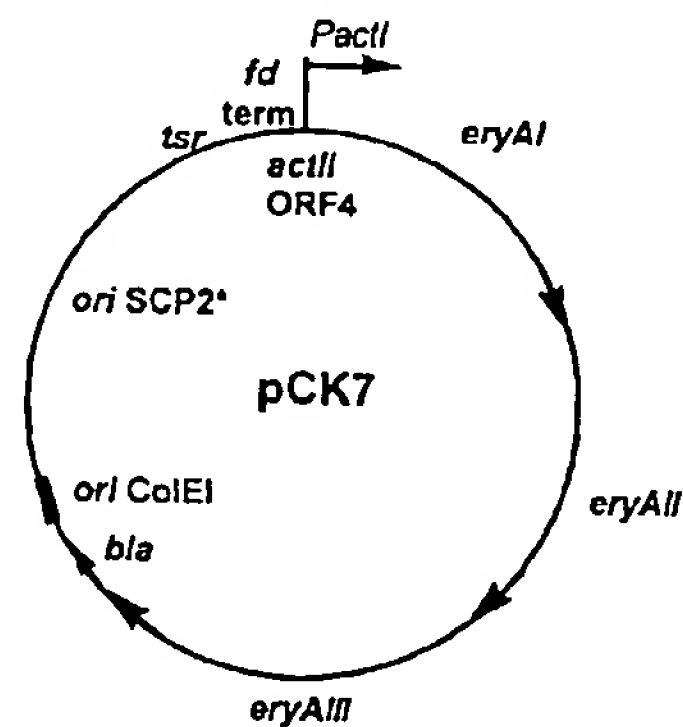
40 25 Cultures of *Myxococcus xanthus* containing the epo genes are grown in a number of media and examined for production of epothilones. If the levels of production of epothilones (in particular B or D) are too low to permit large scale fermentation, the
45 *M. xanthus*-producing clones are subjected to media development and strain improvement, as described below for enhancing production in *Streptomyces*.

30 Example 4

Construction of a *Streptomyces* Expression Vector

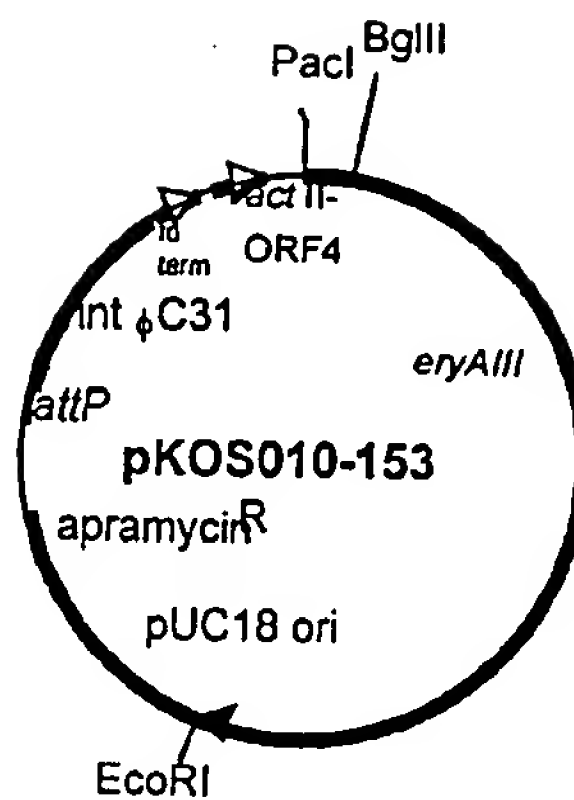
50 The present invention provides recombinant expression vectors for the heterologous expression of modular polyketide synthase genes in *Streptomyces* hosts.

These vectors include expression vectors that employ the *actI* promoter that is regulated by the gene *actII* ORF4 to allow regulated expression at high levels when growing cells enter stationary phase. Among the vectors available are plasmids pRM1 and pRM5, and derivatives thereof such as pCK7, which are stable, low copy plasmids that carry the marker for thiostrepton resistance in actinomycetes. Such plasmids can accommodate large inserts of cloned DNA and have been used for the expression of the DEBS PKS in *S. coelicolor* and *S. lividans*, the picromycin PKS genes in *S. lividans*, and the oleandomycin PKS genes in *S. lividans*. See U.S. Patent No. 5,712,146. Those of skill in the art recognize that *S. lividans* does not make the tRNA that recognizes the TTA codon for leucine until late-stage growth and that if production of a protein is desired earlier, then appropriate codon modifications can be made.



Plasmid pCK7

Another vector is a derivative of plasmid pSET152 and comprises the *actII* ORF4-*PactI* expression system but carries the selectable marker for apramycin resistance. These vectors contain the *attP* site and integrase gene of the actinophage *phiC31* and do not replicate autonomously in *Streptomyces* hosts but integrate by site specific recombination into the chromosome at the attachment site for *phiC31* after introduction into the cell. Derivatives of pCK7 and pSET152 have been used together for the heterologous production of a polyketide, with different PKS genes expressed from each plasmid. See U.S. patent application Serial No. 60/129,731, filed 16 Apr. 1999, incorporated herein by reference.



Plasmid pKOS010-153, a pSET152 Derivative

The need to develop expression vectors for the epothilone PKS that function in *Streptomyces* is significant. The epothilone compounds are currently produced in the slow growing, genetically intractable host *Sorangium cellulosum* or are made synthetically. The streptomycetes, bacteria that produce more than 70% of all known antibiotics and important complex polyketides, are excellent hosts for production of epothilones and epothilone derivatives. *S. lividans* and *S. coelicolor* have been developed for the expression of heterologous PKS systems. These organisms can stably maintain cloned heterologous PKS genes, express them at high levels under controlled conditions, and modify the corresponding PKS proteins (e.g. phosphopantetheinylation) so that they are capable of production of the polyketide they encode. Furthermore, these hosts contain the necessary pathways to produce the substrates required for polyketide synthesis, e.g. malonyl CoA and methylmalonyl CoA. A wide variety of cloning and expression vectors are available for these hosts, as are methods for the introduction and stable maintenance of large segments of foreign DNA. Relative to the slow growing *Sorangium* host, *S. lividans* and *S. coelicolor* grow well on a number of media and have been adapted for high level production of polyketides in fermentors. A number of approaches are available for yield improvements, including rational approaches to increase expression rates, increase precursor supply, etc. Empirical methods to increase the titers of the polyketides, long since proven effective for numerous other polyketides produced in streptomycetes, can also be employed for the epothilone and epothilone derivative producing host cells of the invention.

5 To produce epothilones by fermentation in a heterologous *Streptomyces* host, the epothilone PKS (including the NRPS module) genes are cloned in two segments in derivatives of pCK7 (loading domain through module 6) and pKOS010-153 (modules 7
10 through 9). The two plasmids are introduced into *S. lividans* employing selection for thioestrepton and apramycin resistance. In this arrangement, the pCK7 derivative replicates autonomously whereas the pKOS010-153 derivative is integrated in the chromosome. In
15 both vectors, expression of the epothilone genes is from the actI promoter resident within the plasmid.

To facilitate the cloning, the two epothilone PKS encoding segments (one for the
10 loading domain through module six and one for modules seven through nine) were cloned as translational fusions with the N-terminal segment of the KS domain of module 5 of the ery PKS. High level expression has been demonstrated from this promoter employing KS5
20 as the first translated sequence, see Jacobsen *et al.*, 1998, Biochemistry 37: 4928-4934, incorporated herein by reference. A convenient BsaBI site is contained within the DNA
25 segment encoding the amino acid sequence EPIAV that is highly conserved in many KS domains including the KS-encoding regions of *epoA* and of module 7 in *epoE*.

The expression vector for the loading domain and modules one through six of the
30 epothilone PKS was designated pKOS039-124, and the expression vector for modules seven through nine was designated pKOS039-126. Those of skill in the art will recognize that other vectors and vector components can be used to make equivalent vectors. Because
20 preferred expression vectors of the invention, described below and derived from pKOS039-124 and pKOS039-126, have been deposited under the terms of the Budapest
35 Treaty, only a summary of the construction of plasmids pKOS039-124 and pKOS039-126 is provided below.

25 The eryKS5 linker coding sequences were cloned as an ~0.4 kb PacI-BglII restriction fragment from plasmid pKOS10-153 into pKOS039-98 to construct plasmid pKOS039-117. The coding sequences for the eryKS5 linker were linked to those for the
40 epothilone loading domain by inserting the ~8.7 kb EcoRI-XbaI restriction fragment from cosmid pKOS35-70.1A2 into EcoRI-XbaI digested plasmid pLitmus28. The ~3.4 kb of
45 BsaBI-NotI and ~3.7 kb NotI-HindIII restriction fragments from the resulting plasmid were inserted into BsaBI-HindIII digested plasmid pKOS039-117 to construct plasmid pKOS039-120. The ~7 kb PacI-XbaI restriction fragment of plasmid pKOS039-120 was
30 inserted into plasmid pKAO18' to construct plasmid pKOS039-123. The final pKOS039-
50

124 expression vector was constructed by ligating the ~34 kb XbaI-AvrII restriction fragment of cosmid pKOS35-70.1A2 with the ~21.1 kb AvrII-XbaI restriction fragment of pKOS039-123.

The plasmid pKOS039-126 expression vector was constructed as follows. First the coding sequences for module 7 were linked from cosmids pKOS35-70.4 and pKOS35-79.85 by cloning the ~6.9 kb BglII-NotI restriction fragment of pKOS35-70.4 and the ~5.9 kb NotI-HindIII restriction fragment of pKOS35-79.85 into BglII-HindIII digested plasmid pLitmus28 to construct plasmid pKOS039-119. The ~12 kb NdeI-NheI restriction fragment of cosmid pKOS35-79.85 was cloned into NdeI-XbaI digested plasmid pKOS039-119 to construct plasmid pKOS039-122.

To fuse the eryKS5 linker coding sequences with the coding sequences for module 7, the ~1 kb BsaBI-BglII restriction fragment derived from cosmid pKOS35-70.4 was cloned into BsaBI-BclI digested plasmid pKOS039-117 to construct plasmid pKOS039-121. The ~21.5 kb AvrII restriction fragment from plasmid pKOS039-122 was cloned into AvrII-XbaI digested plasmid pKOS039-121 to construct plasmid pKOS039-125. The ~21.8 kb PacI-EcoRI restriction fragment of plasmid pKOS039-125 was ligated with the ~9 kb PacI-EcoRI restriction fragment of plasmid pKOS039-44 to construct pKOS039-126.

Plasmids pKOS039-124 and pKOS126 were introduced into *S. lividans* K4-114 sequentially employing selection for the corresponding drug resistance marker. Because plasmid pKOS039-126 does not replicate autonomously in streptomycetes, the selection is for cells in which the plasmid has integrated in the chromosome by site-specific recombination at the attB site of phiC31. Because the plasmid stably integrates, continued selection for apramycin resistance is not required. Selection can be maintained if desired. The presence of thiostrepton in the medium is maintained to ensure continued selection for plasmid pKOS039-124. Plasmids pKOS039-124 and pKOS039-126 were transformed into *Streptomyces lividans* K4-114, and transformants containing the plasmids were cultured and tested for production of epothilones. Initial tests did not indicate the presence of an epothilone.

To improve production of epothilones from these vectors, the eryKS5 linker sequences were replaced by epothilone PKS gene coding sequences, and the vectors were introduced into *Streptomyces coelicolor* CH999. To amplify by PCR coding sequences from the *epoA* gene coding sequence, two oligonucleotides primers were used:

5 N39-73, 5'-GCTTAATTAAGGAGGACACATATGCCCGTCGTGGCGGATCGTCC-3'; and
N39-74, 5'-GCGGATCCTCGAATCACCGCCAATATC-3'.

10 The template DNA was derived from cosmid pKOS35-70.8A3. The ~0.8 kb PCR product
was digested with restriction enzymes *PacI* and *BamHI* and then ligated with the ~2.4 kb
5 *BamHI*-*NotI* and the ~6.4 kb *PacI*-*NotI* restriction fragments of plasmid pKOS039-120 to
construct plasmid pKOS039-136. To make the expression vector for the *epoA*, *epoB*,
epoC, and *epoD* genes, the ~5 kb *PacI*-*AvrII* restriction fragment of plasmid pKOS039-
15 136 was ligated with the ~50 kb *PacI*-*AvrII* restriction fragment of plasmid pKOS039-124
to construct the expression plasmid pKOS039-124R. Plasmid pKOS039-124R has been
10 deposited with the ATCC under the terms of the Budapest Treaty and is available under
accession number _____.

20 To amplify by PCR sequences from the *epoE* gene coding sequence, two
oligonucleotide primers were used:

N39-67A, 5'-GCTTAATTAAGGAGGACACATATGACCGACCGAGAAGGCCAGCTC-CTGGA-3', and
15 N39-68, 5'-GGACCTAGGCGGGATGCCGGCGTCT-3'.

25 The template DNA was derived from cosmid pKOS35-70.1A2. The ~0.4 kb
amplification product was digested with restriction enzymes *PacI* and *AvrII* and ligated
with either the ~29.5 kb *PacI*-*AvrII* restriction fragment of plasmid pKOS039-126 or the
30 ~23.8 kb *PacI*-*AvrII* restriction fragment of plasmid pKOS039-125 to construct plasmid
pKOS039-126R or plasmid pKOS039-125R, respectively. Plasmid pKOS039-126R was
20 deposited with the ATCC under the terms of the Budapest Treaty and is available under
accession number _____.

35 The plasmid pair pKOS039-124R and pKOS039-126R (as well as the plasmid pair
pKOS039-124 and pKOS039-126) contain the full complement of *epoA*, *epoB*, *epoC*,
25 *epoD*, *epoE*, *epoF*, *epoK*, and *epoL* genes. The latter two genes are present on plasmid
pKOS039-126R (as well as plasmid pKOS039-126); however, to ensure that these genes
40 were expressed at high levels, another expression vector of the invention, plasmid
pKOS039-141 (Figure 8), was constructed in which the *epoK* and *epoL* genes were placed
under the control of the *ermE** promoter.

45 30 The *epoK* gene sequences were amplified by PCR using the oligonucleotide
primers:

N39-69, 5'-AGGCATGCATATGACCCAGGAGCAAGCGAATCAGAGTG-3'; and
50 N39-70, 5'-CCAAGCTTTATCCAGCTTTGGAGGGCTTCAAG-3'.

5 The *epoL* gene sequences were amplified by PCR using the oligonucleotide primers:

N39-71A, 5'-GTAAGCTTAGGAGGACACATATGATGCAACTCGCGCGGGTG-3'; and
10 N39-72, 5'-GCCTGCAGGCTCAGGCTTGCCGAGAGCGT-3'.

5 The template DNA for the amplifications was derived from cosmid pKOS35-79.85. The PCR products were subcloned into PCR-script for sequence analysis. Then, the *epoK* and *epoL* genes were isolated from the clones as NdeI-HindIII and HindIII-EcoRI restriction fragments, respectively, and ligated with the ~6 kb NdeI-EcoRI restriction
15 fragment of plasmid pKOS039-134B, which contains the *ermE** promoter, to construct plasmid pKOS039-140. The ~2.4 kb NheI-PstI restriction fragment of plasmid pKOS039-140 was cloned into XbaI-PstI digested plasmid pSAM-Hyg, a plasmid pSAM2 derivative
20 containing a hygromycin resistance conferring gene, to construct plasmid pKOS039-141.

Another variant of plasmid pKOS039-126R was constructed to provide the *epoE* and *epoF* genes on an expression vector without the *epoK* and *epoL* genes. This plasmid,
25 pKOS045-12 (Figure 9), was constructed as follows. Plasmid pXH106 (described in J. Bact., 1991, 173: 5573-5577, incorporated herein by reference) was digested with restriction enzymes StuI and BamHI, and the ~2.8 kb restriction fragment containing the *xylE* and hygromycin resistance conferring genes was isolated and cloned into EcoRV-
30 BglII digested plasmid pLitmus28. The ~2.8 kb NcoI-AvrII restriction fragment of the resulting plasmid was ligated to the ~18 kb PacI-BspIII restriction fragment of plasmid pKOS039-125R and the ~9 kb SpeI-PacI restriction fragment of plasmid pKOS039-42 to
20 construct plasmid pKOS045-12.

To construct an expression vector that comprised only the *epoL* gene, plasmid pKOS039-141 was partially digested with restriction enzyme NdeI, the ~9 kb NdeI
25 restriction fragment was isolated, and the fragment then circularized by ligation to yield plasmid pKOS039-150.
40

The various expression vectors described above were then transformed into *Streptomyces coelicolor* CH999 and *S. lividans* K4-114 in a variety of combinations, the
45 transformed host cells fermented on plates and in liquid culture (R5 medium, which is identical to R2YE medium without agar). Typical fermentation conditions follow. First, a
30 seed culture of about 5 mL containing 50 µg/L thiostrepton was inoculated and grown at 30°C for two days. Then, about 1 to 2 mL of the seed culture was used to inoculate a
50 production culture of about 50 mL containing 50 µg/L thiostrepton and 1 mM cysteine,

5 and the production culture was grown at 30°C for 5 days. Also, the seed culture was used to prepare plates of cells (the plates contained the same media as the production culture with 10 mM propionate), which were grown at 30°C for nine days.

10 Certain of the *Streptomyces coelicolor* cultures and culture broths were analyzed for production of epothilones. The liquid cultures were extracted with three times with equal volumes of ethyl acetate, the organic extracts combined and evaporated, and the residue dissolved in acetonitrile for LC/MS analysis. The agar plate media was chopped and extracted twice with equal volumes of acetone, and the acetone extracts were combined and evaporated to an aqueous slurry, which was extracted three times with equal
15 volumes of ethyl acetate. The organic extracts were combined and evaporated, and the residue dissolved in acetonitrile for LC/MS analysis.

20 Production of epothilones was assessed using LC-mass spectrometry. The output flow from the UV detector of an analytical HPLC was split equally between a Perkin-Elmer/Sciex API100LC mass spectrometer and an Alltech 500 evaporative light scattering
25 detector. Samples were injected onto a 4.6 x 150 mm reversed phase HPLC column (MetaChem 5 m ODS-3 Inertsil) equilibrated in water with a flow rate of 1.0 mL/min. UV detection was set at 250 nm. Sample components were separated using H₂O for 1 minute, then a linear gradient from 0 to 100% acetonitrile over 10 minutes. Under these
30 conditions, epothilone A elutes at 10.2 minutes and epothilone B elutes at 10.5 minutes. The identity of these compounds was confirmed by the mass spectra obtained using an atmospheric chemical ionization source with orifice and ring voltages set at 75 V and 300
35 V, respectively, and a mass resolution of 0.1 amu. Under these conditions, epothilone A shows [M+H] at 494.4 amu, with observed fragments at 476.4, 318.3, and 306.4 amu. Epothilone B shows [M+H] at 508.4 amu, with observed fragments at 490.4, 320.3, and
40 302.4 amu.

45 Transformants containing the vector pairs pKOS039-124R and pKOS039-126R or pKOS039-124 and pKOS039-126R produced detectable amounts of epothilones A and B. Transformants containing these plasmid pairs and the additional plasmid pKOS039-141 produced similar amounts of epothilones A and B, indicating that the additional copies of
30 the *epoK* and *epoL* genes were not required for production under the test conditions employed. Thus, these transformants produced epothilones A and B when recombinant *epoA*, *epoB*, *epoC*, *epoD*, *epoE*, *epoF*, *epoK*, and *epoL* genes were present. In some
50

5 cultures, it was observed that the absence of propionate increased the proportion of
epothilone B to epothilone A.

10 Transformants containing the plasmid pair pKOS039-124R and pKOS045-12
produced epothilones C and D, as did transformants containing this plasmid pair and the
5 additional plasmid pKOS039-150. These results showed that the *epoL* gene was not
required under the test conditions employed to form the C-12-C-13 double bond. These
results indicate that either the epothilone PKS gene alone is able to form the double bond
15 or that *Streptomyces coelicolor* expresses a gene product able to convert epothilones G and
H to epothilones C and D. Thus, these transformants produced epothilones C and D when
10 recombinant *epoA*, *epoB*, *epoC*, *epoD*, *epoE*, and *epoF* genes were present.

20 The heterologous expression of the epothilone PKS described herein is believed to
represent the recombinant expression of the largest proteins and active enzyme complex
that have ever been expressed in a recombinant host cell. The epothilone producing
Streptomyces coelicolor transformants exhibited growth characteristics indicating that
25 15 either the epothilone PKS genes, or their products, or the epothilones inhibited cell growth
or were somewhat toxic to the cells. Any such inhibition or toxicity could be due to
accumulation of the epothilones in the cell, and it is believed that the native *Sorangium*
producer cells may contain transporter proteins that in effect pump epothilones out of the
30 cell. Such transporter genes are believed to be included among the ORFs located
20 downstream of the *epoK* gene and described above. Thus, the present invention provides
Streptomyces and other host cells that include recombinant genes that encode the products
35 of one or more, including all, of the ORFs in this region.

For example, each ORF can be cloned behind the *ermE** promoter, see Stassi *et*
al., 1998, Appl. Microbiol. Biotechnol. 49: 725-731, incorporated herein by reference, in a
25 pSAM2-based plasmid that can integrate into the chromosome of *Streptomyces coelicolor*
40 and *S. lividans* at a site distinct from attB of phage phiC31, see Smokvina *et al.*, 1990,
Gene 94: 53-59, incorporated herein by reference. A pSAM2-based vector carrying the
gene for hygromycin resistance is modified to carry the *ermE** promoter along with
45 additional cloning sites. Each ORF downstream is PCR cloned into the vector which is
30 then introduced into the host cell (also containing pKOS039-124R and pKOS039-126R or
other expression vectors of the invention) employing hygromycin selection. Clones
50 carrying each individual gene downstream from *epoK* are analyzed for increased
production of epothilones.

5 Additional fermentation and strain improvement efforts can be conducted as
illustrated by the following. The levels of expression of the PKS genes in the various
constructs can be measured by assaying the levels of the corresponding mRNAs (by
10 quantitative RT PCR) relative to the levels of another heterologous PKS mRNA (e.g.
5 picromycin) produced from genes cloned in similar expression vectors in the same host. If
one of the epothilone transcripts is underproduced, experiments to enhance its production
by cloning the corresponding DNA segment in a different expression vector are
15 conducted. For example, multiple copies of any one or more of the epothilone PKS genes
can be introduced into a cell if one or more gene products are rate limiting for
10 biosynthesis. If the basis for low level production is not related to low level PKS gene
expression (at the RNA level), an empirical mutagenesis and screening approach that is
20 the backbone of yield improvement of every commercially important fermentation product
is undertaken. Spores are subjected to UV, X-ray or chemical mutagens, and individual
survivors are plated and picked and tested for the level of compound produced in small
25 scale fermentations. Although this process can be automated, one can examine several
thousand isolates for quantifiable epothilone production using the susceptible fungus
15 *Mucor hiemalis* as a test organism.

30 Another method to increase the yield of epothilones produced is to change the KS^Y
domain of the loading domain of the epothilone PKS to a KS^Q domain. Such altered
20 loading domains can be constructed in any of a variety of ways, but one illustrative
method follows. Plasmid pKOS39-124R of the invention can be conveniently used as a
35 starting material. To amplify DNA fragments useful in the construction, four
oligonucleotide primers are employed:
N39-83: 5'-CCGGTATCCACCGCGACACACGGC-3',
25 N39-84: 5'-GCCAGTCGTCCTCGCTCGTGGCCGTTC-3',
40 and N39-73 and N39-74, which have been described above. The PCR fragment generated
with N39-73 and N39-83 and the PCR fragment generated with N39-74 and N39-84 are
treated with restriction enzymes PacI and BamHI, respectively, and ligated with the ~3.1
45 kb PacI-BamHI fragment of plasmid pKOS39-120 to construct plasmid pKOS039-148.
30 The ~0.8 kb PacI-BamHI restriction fragment of plasmid pKOS039-148 (comprising the
two PCR amplification products) is ligated with the ~2.4 kb BamHI-NotI restriction
fragment and the ~6.4 kb PacI-NotI restriction fragment of plasmid pKOS39-120 to
50 construct pKOS39-136Q. The ~5 kb PacI-AvrII restriction fragment of plasmid pKOS039-

5 136Q is ligated to the ~50 kb PacI-AvrII restriction fragment of plasmid pKOS039-124 to construct plasmid pKOS39-124Q. Plasmids pKOS039-124Q and pKOS039-126R are then transformed into *Streptomyces coelicolor* CH999 for epothilone production.

10 The *epoA* through *epoF*, optionally with *epoK* or with *epoK* plus *epoL*, genes
5 cloned and expressed are sufficient for the synthesis of epothilone compounds, and the distribution of the C-12 H to C-12 methyl congeners appears to be similar to that seen in the natural host (A:B::2:1). This ratio reflects that the AT domain of module 4 more
15 closely resembles that of the malonyl rather than methylmalonyl specifying AT consensus domains. Thus, epothilones D and B are produced at lower quantities than their C-12
10 unmethylated counterparts C and A. The invention provides PKS genes that produce epothilone D and/or B exclusively. Specifically, methylmalonyl CoA specifying AT
20 domains from a number of sources (e.g. the narbonolide PKS, the rapamycin PKS, and others listed above) can be used to replace the naturally occurring at domain in module 4. The exchange is performed by direct cloning of the incoming DNA into the appropriate
25 15 site in the epothilone PKS encoding DNA segment or by gene replacement through homologous recombination.

For gene replacement through homologous recombination, the donor sequence to
30 be exchanged is placed in a delivery vector between segments of at least 1 kb in length that flank the AT domain of epo module 4 encoding DNA. Crossovers in the homologous
20 regions result in the exchange of the epo AT4 domain with that on the delivery vector. Because pKOS039-124 and pKOS039-124R contain AT4 coding sequences, they can be
35 used as the host DNA for replacement. The adjacent DNA segments are cloned in one of a number of *E. coli* plasmids that are temperature sensitive for replication. The heterologous
AT domains can be cloned in these plasmids in the correct orientation between the
25 homologous regions as cassettes enabling the ability to perform several AT exchanges simultaneously. The reconstructed plasmid (pKOS039-124* or pKOS039-124R*) is tested
40 for ability to direct the synthesis of epothilone B and/or by introducing it along with pKOS039-126 or pKOS039-126R in *Streptomyces coelicolor* and/or *S. lividans*.

45 Because the titers of the polyketide can vary from strain to strain carrying the
30 different gene replacements, the invention provides a number of heterologous methylmalonyl CoA specifying AT domains to ensure that production of epothilone D at
titers equivalent to that of the C and D mixture produced in the *Streptomyces coelicolor*
50 host described above. In addition, larger segments of the donor genes can be used for the

5 replacements, including, in addition to the AT domain, adjacent upstream and downstream
 sequences that correspond to an entire module. If an entire module is used for the
 replacement, the KS, methylmalonyl AT, DH, KR, ACP – encoding DNA segment can be
 10 obtained from for example and without limitation the DNA encoding the tenth module of
 5 the rapamycin PKS, or the first or fifth modules of the FK-520 PKS.

Example 5

Heterologous Expression of EpoK and Conversion of Epothilone D to Epothilone B

15 This Example describes the construction of *E. coli* expression vectors for *epoK*.
 10 The *epoK* gene product was expressed in *E. coli* as a fusion protein with a polyhistidine
 tag (his tag). The fusion protein was purified and used to convert epothilone D to
 20 epothilone B.

Plasmids were constructed to encode fusion proteins composed of six histidine
 residues fused to either the amino or carboxy terminus of EpoK. The following oligos
 25 were used to construct the plasmids:

55-101.a-1:

5'-AAAAACATATGCACCACCACCACCACATGACACAGGAGCAAGCGAAT-CAGAGTGAG-3',

55-101.b:

5'-AAAAAGGATCCTTAATCCAGCTTTGGAGGGCTT-3',

20 55-101.c:

5'-AAAAACATATGACACAGGAGCAAGCGAAT-3', and

55-101.d:

35 5'-AAAAAGCATCCTTAGTGGTGGTGGTGGTGGTGTCCAGCTTTGGAGGGCTTC-AAGATGAC-3'.

The plasmid encoding the amino terminal his tag fusion protein, pKOS55-121, was
 25 constructed using primers 55-101.a-1 and 55-101.b, and the one encoding the carboxy
 terminal his tag, pKOS55-129, was constructed using primers 55-101.c and 55-101.d in
 40 PCR reactions containing pKOS35-83.5 as the template DNA. Plasmid pKOS35-83.5
 contains the ~5 kb NotI fragment comprising the *epoK* gene ligated into pBluescriptSKII+
 (Stratagene). The PCR products were cleaved with restriction enzymes BamHI and NdeI
 30 and ligated into the BamHI and NdeI sites of pET22b (Invitrogen). Both plasmids were
 45 sequenced to verify that no mutations were introduced during the PCR amplification.
 Protein gels were run as known in the art.

Purification of EpoK was performed as follows. Plasmids pKOS55-121 and
 50 pKOS55-129 were transformed into BL21(DE3) containing the groELS expressing

5 plasmid pREP4-groELS (Caspers *et al.*, 1994, Cellular and Molecular Biology 40(5):
635-644). The strains were inoculated into 250 mL of M9 medium supplemented with 2
10 mM MgSO₄, 1% glucose, 20 mg thiamin, 5 mg FeCl₂, 4 mg CaCl₂ and 50 mg levulinic
acid. The cultures were grown to an OD₆₀₀ between 0.4 and 0.6, at which point IPTG was
5 added to 1 mM, and the cultures were allowed to grow for an additional two hours. The
cells were harvested and frozen at -80°C. The frozen cells were resuspended in 10 ml of
buffer 1 (5 mM imidazole, 500 mM NaCl, and 45 mM Tris pH 7.6) and were lysed by
15 sonicating three times for 15 seconds each on setting 8. The cellular debris was pelleted by
spinning in an SS-34 rotor at 16,000 rpm for 30 minutes. The supernatant was removed
10 and spun again at 16,000 rpm for 30 minutes. The supernatant was loaded onto a 5 mL
nickel column (Novagen), after which the column was washed with 50 mL of buffer 1
20 (Novagen). EpoK was eluted with a gradient from 5 mM to 1M imidazole. Fractions
containing EpoK were pooled and dialyzed twice against 1 L of dialysis buffer (45 mM
Tris pH7.6, 0.2 mM DTT, 0.1 mM EDTA, and 20% glycerol). Aliquots were frozen in
25 liquid nitrogen and stored at -80°C. The protein preparations were greater than 90% pure.

The EpoK assay was performed as follows (See Betlach *et al.*, *Biochem* (1998)
37:14937, incorporated herein by reference). Briefly, reactions consisted of 50 mM Tris
30 (pH7.5), 21 μM spinach ferredoxin, 0.132 units of spinach ferredoxin: NADP⁺
oxidoreductase, 0.8 units of glucose-6-phosphate dehydrogenase, 1.4 mM NADP, and 7.1
20 mM glucose-6-phosphate, 100 μM or 200 μM epothilone D (a generous gift of S.
Danishefsky), and 1.7 μM amino terminal his tagged EpoK or 1.6 μM carboxy terminal
35 his tagged EpoK in a 100 μL volume. The reactions were incubated at 30°C for 67
minutes and stopped by heating at 90°C for 2 minutes. The insoluble material was
removed by centrifugation, and 50 μL of the supernatant were analyzed by LC/MS. HPLC
25 conditions: Metachem 5 μ ODS-3 Inertsil (4.6 X 150 mm); 80% H₂O for 1 min, then to
40 100% MeCN over 10 min at 1 mL/min, with UV (λ_{max} =250 nm), ELSD, and MS
detection. Under these conditions, epothilone D eluted at 11.6 min and cpothilone B at 9.3
min. the LC/MS spectra were obtained using an atmosphere pressure chemical ionization
45 source with orifice and ring voltages set at 20 V and 250 V, respectively, at a mass
30 resolution of 1 amu. Under these conditions, epothilone E shows an [M+H] at *m/z* 493,
with observed fragments at 405 and 304. Epothilone B shows an [M+H] at *m/z* 509, with
observed fragments at 491 and 320.

5 The reactions containing EpoK and epothilone D contained a compound absent in
the control that displayed the same retention time, molecular weight, and mass
10 fragmentation pattern as pure epothilone B. With an epothilone D concentration of 100
μM, the amino and the carboxy terminal his tagged EpoK was able to convert 82% and
5 58% to epothilone B, respectively. In the presence of 200 μM, conversion was 44% and
21%, respectively. These results demonstrate that EpoK can convert epothilone D to
15 epothilone B.

10 Example 6

Modified Epothilones from Chemobiosynthesis

20 This Example describes a series of thioesters provided by the invention for
production of epothilone derivatives via chemobiosynthesis. The DNA sequence of the
biosynthetic gene cluster for epothilone from *Sorangium cellulosum* indicates that priming
25 of the PKS involves a mixture of polyketide and amino acid components. Priming
15 involves loading of the PKS-like portion of the loading domain with malonyl CoA
followed by decarboxylation and loading of the module one NRPS with cysteine, then
condensation to form enzyme-bound N-acetylcysteine. Cyclization to form a thiazoline is
30 followed by oxidation to form enzyme bound 2-methylthiazole-4-carboxylate, the product
of the loading domain and NRPS. Subsequent condensation with methylmalonyl CoA by
20 the ketosynthase of module 2 provides the substrate for module, as shown in the following
35 diagram.

5

10

15

20

25

30

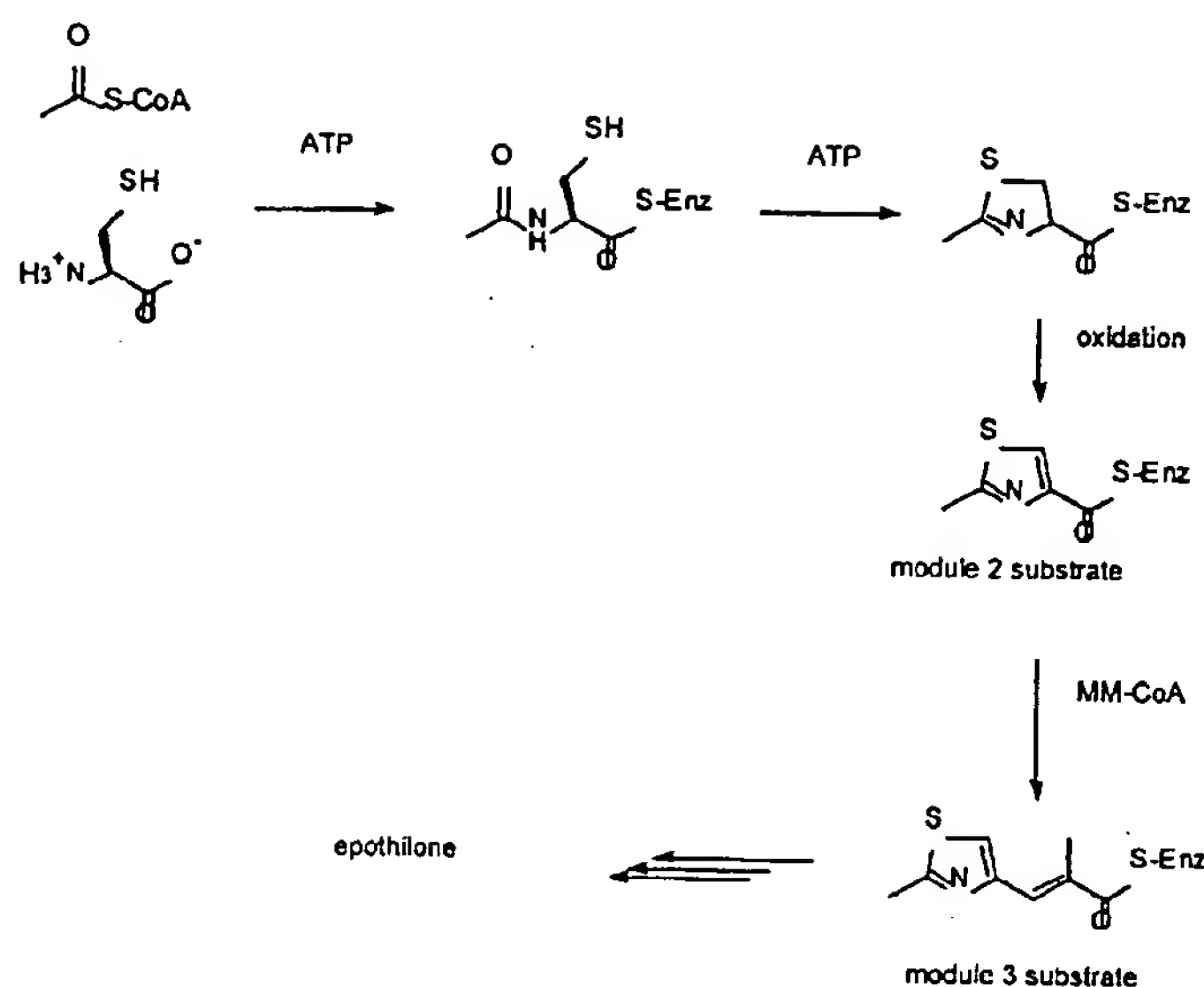
35

40

45

50

55

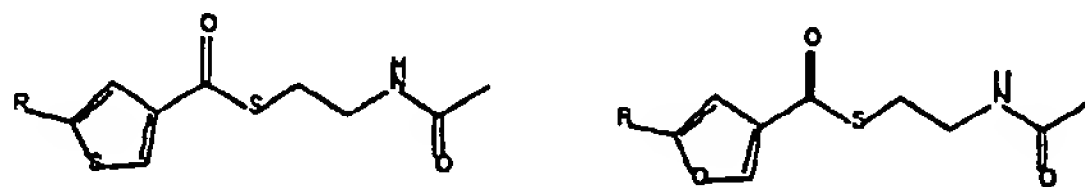


The present invention provides methods and reagents for chemobiosynthesis to produce epothilone derivatives in a manner similar to that described to make 6-dEB and erythromycin analogs in PCT Pat. Pub. Nos. 99/03986 and 97/02358. Two types of feeding substrates are provided: analogs of the NRPS product, and analogs of the module 3 substrate. The module 2 substrates are used with PKS enzymes with a mutated NRPS-like domain, and the module 3 substrates are used with PKS enzymes with a mutated KS domain in module 2.

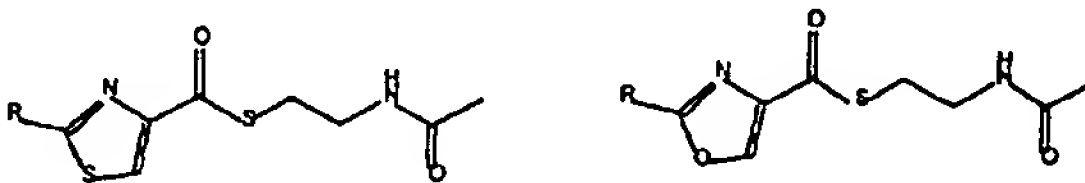
5

The following illustrate module 2 substrates (as N-acetyl cysteamine thioesters) for use as substrates for epothilone PKS with modified inactivated NRPS:

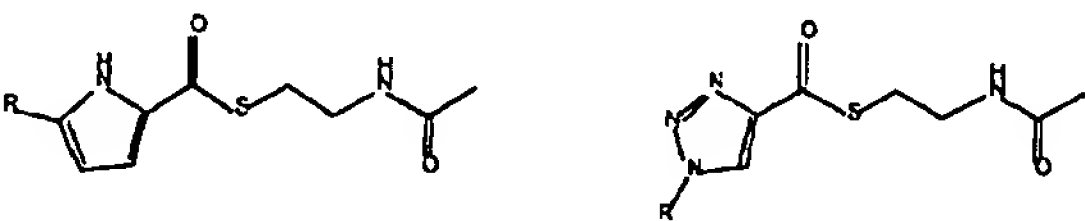
10



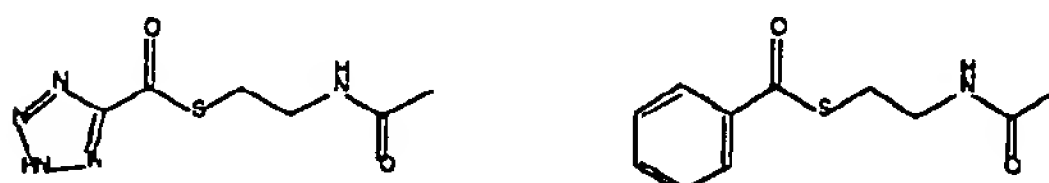
15



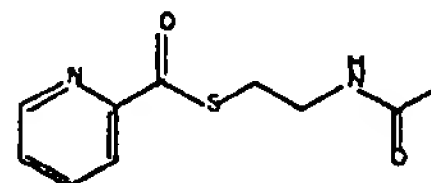
20



25



30



The module 2 substrates are prepared by activation of the corresponding carboxylic acid and treatment with N-acetylcysteamine. Activation methods include formation of the acid chloride, formation of a mixed anhydride, or reaction with a condensing reagent such as a carbodiimide.

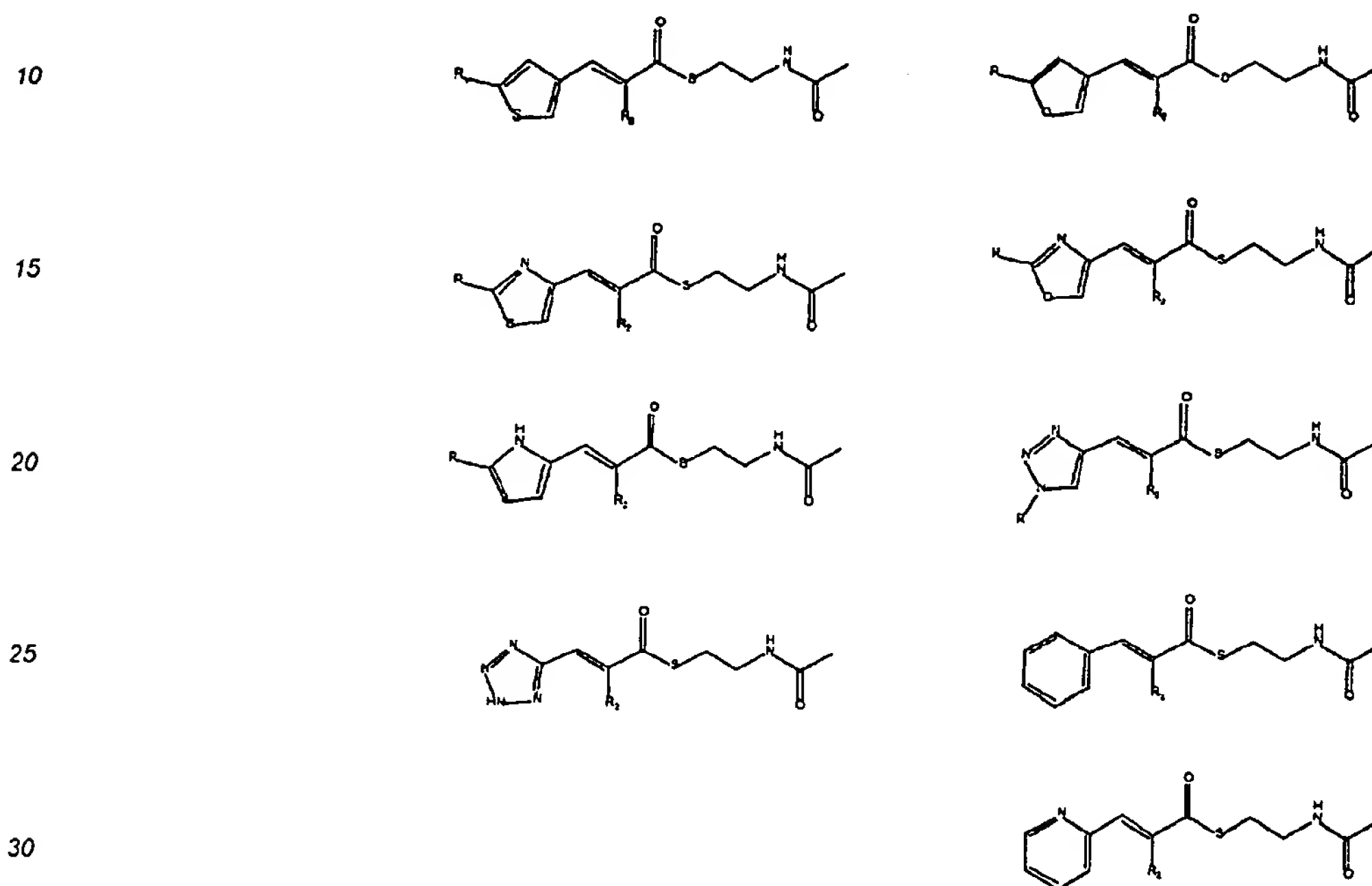
40

45

50

55

Exemplary module 3 substrates, also as NAc thioesters for use as substrates for epothilone PKS with KS2 knockout are:



These compounds are prepared in a three-step process. First, the appropriate aldehyde is treated with a Wittig reagent or equivalent to form the substituted acrylic ester. The ester is saponified to the acid, which is then activated and treated with N-acetylcysteamine.

Illustrative reaction schemes for making module 2 and module 3 substrates follow. Additional compounds suitable for making starting materials for polyketide synthesis by the epothilone PKS are shown in Figure 2 as carboxylic acids (or aldehydes that can be converted to carboxylic acids) that are converted to the N-acetylcysteamides for supplying to the host cells of the invention.

A. Thiophene-3-carboxylate N-acetylcysteamine thioester

A solution of thiophene-3-carboxylic acid (128 mg) in 2 mL of dry tetrahydrofuran under inert atmosphere was treated with triethylamine (0.25 mL) and diphenylphosphoryl azide (0.50 mL). After 1 hour, N-acetylcysteamine (0.25 mL) was added, and the reaction

5 was allowed to proceed for 12 hours. The mixture was poured into water and extracted
three times with equal volumes of ethyl acetate. The organic extracts were combined,
washed sequentially with water, 1 N HCl, sat. CuSO₄, and brine, then dried over MgSO₄,
10 filtered, and concentrated under vacuum. Chromatography on SiO₂ using ether followed
5 by ethyl acetate provided pure product, which crystallized upon standing.

15 B. Furan-3-carboxylate N-acetylcysteamine thioester

A solution of furan-3-carboxylic acid (112 mg) in 2 mL of dry tetrahydrofuran
under inert atmosphere was treated with triethylamine (0.25 mL) and diphenylphosphoryl
10 azide (0.50 mL). After 1 hour, N-acetylcysteamine (0.25 mL) was added and the reaction
was allowed to proceed for 12 hours. The mixture was poured into water and extracted
20 three times with equal volumes of ethyl acetate. The organic extracts were combined,
washed sequentially with water, 1 N HCl, sat. CuSO₄, and brine, then dried over MgSO₄,
filtered, and concentrated under vacuum. Chromatography on SiO₂ using ether followed
25 15 by ethyl acetate provided pure product, which crystallized upon standing.

30 C. Pyrrole-2-carboxylate N-acetylcysteamine thioester

A solution of pyrrole-2-carboxylic acid (112 mg) in 2 mL of dry tetrahydrofuran
under inert atmosphere was treated with triethylamine (0.25 mL) and diphenylphosphoryl
20 azide (0.50 mL). After 1 hour, N-acetylcysteamine (0.25 mL) was added and the reaction
was allowed to proceed for 12 hours. The mixture was poured into water and extracted
35 three times with equal volumes of ethyl acetate. The organic extracts were combined,
washed sequentially with water, 1 N HCl, sat. CuSO₄, and brine, then dried over
MgSO₄, filtered, and concentrated under vacuum. Chromatography on SiO₂ using ether
25 followed by ethyl acetate provided pure product, which crystallized upon standing.

40 D. 2-Methyl-3-(3-thienyl)acrylate N-acetylcysteamine thioester

(1) Ethyl 2-methyl-3-(3-thienyl)acrylate: A mixture of thiophene-3-
45 carboxaldehyde (1.12 g) and (carbethoxyethylidene)triphenylphosphorane (4.3 g) in dry
30 tetrahydrofuran (20 mL) was heated at reflux for 16 hours. The mixture was cooled to
ambient temperature and concentrated to dryness under vacuum. The solid residue was
suspended in 1:1 ether/hexane and filtered to remove triphenylphosphine oxide. The
50

5 filtrate was filtered through a pad of SiO₂ using 1:1 ether/hexane to provide the product (1.78 g, 91%) as a pale yellow oil.

10 (2) 2-Methyl-3-(3-thienyl)acrylic acid: The ester from (1) was dissolved in a mixture of methanol (5 mL) and 8 N KOH (5 mL) and heated at reflux for 30 minutes. The mixture was cooled to ambient temperature, diluted with water, and washed twice with
5 ether. The aqueous phase was acidified using 1N HCl then extracted 3 times with equal volumes of ether. The organic extracts were combined, dried with MgSO₄, filtered, and concentrated to dryness under vacuum. Crystallization from 2:1 hexane/ether provided the
15 product as colorless needles.

10 (3) 2-Methyl-3-(3-thienyl)acrylate N-acetylcysteamine thioester: A solution of 2-Methyl-3-(3-thienyl)acrylic acid (168 mg) in 2 mL of dry tetrahydrofuran under inert
20 atmosphere was treated with triethylamine (0.56 mL) and diphenylphosphoryl azide (0.45 mL). After 15 minutes, N-acetylcysteamine (0.15 mL) is added and the reaction is allowed to proceed for 4 hours. The mixture is poured into water and extracted three times with
25 equal volumes of ethyl acetate. The organic extracts are combined, washed sequentially with water, 1 N HCl, sat. CuSO₄, and brine, then dried over MgSO₄, filtered, and concentrated under vacuum. Chromatography on SiO₂ using ethyl acetate provided pure
30 product, which crystallized upon standing.

35 The above compounds are supplied to cultures of host cells containing a recombinant epothilone PKS of the invention in which either the NRPS or the KS domain of module 2 as appropriate has been inactivated by mutation to prepare the corresponding epothilone derivative of the invention.

Example 7

25 Producing Epothilones and Epothilone Derivatives in *Sorangium cellulosum* SMP44

40 The present invention provides a variety of recombinant *Sorangium cellulosum* host cells that produce less complex mixtures of epothilones than the naturally occurring epothilone producers as well as host cells that produce epothilone derivatives. This
45 Example illustrates the construction of such strains by describing how to make a strain that produce only epothilones C and D without epothilones A and B. To construct this strain,
30 an inactivating mutation is made in *epoK*. Using plasmid pKOS35-83.5, which contains a NotI fragment harboring the *epoK* gene, the kanamycin and bleomycin resistance markers from Tn5 are ligated into the ScaI site of the *epoK* gene to construct pKOS90-55. The
50

5 orientation of the resistance markers is such that transcription initiated at the kanamycin promoter drives expression of genes immediately downstream of *epoK*. In other words, the mutation should be nonpolar. Next, the origin of conjugative transfer, *oriT*, from RP4 is
10 ligated into pKOS90-55 to create pKOS90-63. This plasmid can be introduced into S17-1 and conjugated into SMP44. The transconjugants are selected on phleomycin plates as previously described. Alternatively, electroporation of the plasmid can be achieved using conditions described above for *Myxococcus xanthus*.

15 Because there are three generalized transducing phages for *Myxococcus xanthus*, one can transfer DNA from *M. xanthus* to SMP44. First, the *epoK* mutation is constructed
20 in *M. xanthus* by linearizing plasmid pKOS90-55 and electroporating into *M. xanthus*. Kanamycin resistant colonies are selected and have a gene replacement of *epoK*. This strain is infected with Mx9, Mx8, Mx4 ts18 hft hrm phages to make phage lysates. These lysates are then individually infected into SMP44 and phleomycin resistant colonies are selected. Once the strain is constructed, standard fermentation procedures, as described
25 below, are employed to produce epothilones C and D.

Prepare a fresh plate of *Sorangium* host cells (dispersed) on S42 medium. S42 medium contains tryptone, 0.5 g/L; MgSO₄, 1.5 g/L; HEPES, 12 g/L; agar, 12 g/L, with
30 deionized water. The pH of S42 medium is set to 7.4 with KOH. To prepare S42 medium, after autoclaving at 121°C for at least 30 minutes, add the following ingredients (per liter):
20 CaCl₂, 1 g; K₂HPO₄, 0.06 g; Fe Citrate, 0.008 g; Glucose, 3.5 g; Ammonium sulfate, 0.5 g; Spent liquid medium, 35 mL; and 200 micrograms/mL of kanamycin is added to prevent contamination. Incubate the culture at 32°C for 4-7 days, or until orange sorangia appear on the surface.

To prepare a seed culture for inoculating agar plates/bioreactor, the following
25 protocol is followed. Scrape off a patch of orange *Sorangium* cells from the agar (about 5 mm²) and transfer to a 250 ml baffie flask with 38 mm silicone foam closures containing 50 ml of Soymeal Medium containing potato starch, 8 g; defatted soybean meal, 2 g; yeast extract, 2 g; Iron (III) sodium salt EDTA, 0.008 g; MgSO₄·7H₂O, 1 g; CaCl₂·2H₂O, 1 g;
45 glucose, 2 g; HEPES buffer, 11.5 g. Use deionized water, and adjust pH to 7.4 with 10% KOH. Add 2-3 drops of antifoam B to prevent foaming. Incubate in a coffin shaker for 4-5
30 days at 30°C and 250 RPM. The culture should appear an orange color. This seed culture can be subcultured repeatedly for scale-up to inoculate in the desired volume of production medium.

5

10

15

20

25

30

35

40

45

50

55

The same preparation can be used with Medium 1 containing (per liter) $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 1 g; yeast extract, 2 g; Soytone, 2 g; FeEDTA , 0.008 g; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g; HEPES, 11.5 g. Adjust pH to 7.4 with 10% KOH, and autoclave at 121°C for 30 minutes. Add 8 ml of 40% glucose after sterilization. Instead of a baffle flask, use a 250 ml coiled spring flask with a foil cover. Include 2-3 drops of antifoam B, and incubate in a coffin shaker for 7 days at 37°C and 250 RPM. Subculture the entire 50 mL into 500 mL of fresh medium in a baffled narrow necked Fernbach flask with a 38 mm silicone foam closure. Include 0.5 ml of antifoam to the culture. Incubate under the same conditions for 2-3 days. Use at least a 10% inoculum for a bioreactor fermentation.

To culture on solid media, the following protocol is used. Prepare agar plates containing (per liter of CNS medium) KNO_3 , 0.5 g; Na_2HPO_4 , 0.25 g; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1 g; FeCl_2 , 0.01 g; HEPES, 2.4 g; Agar, 15 g; and sterile Whatman filter paper. While the agar is not completely solidified, place a sterile disk of filter paper on the surface. When the plate is dry, add just enough of the seed culture to coat the surface evenly (about 1 mL). Spread evenly with a sterile loop or an applicator, and place in a 32°C incubator for 7 days. Harvest plates.

For production in a 5 L bioreactor, the following protocol is used. The fermentation can be conducted in a B. Braun Biostat MD-1 5L bioreactor. Prepare 4 L of production medium (same as the soymeal medium for the seed culture without HEPES buffer). Add 2% (volume to volume) XAD-16 absorption resin, unwashed and untreated, e.g. add 1 mL of XAD per 50 mL of production medium. Use 2.5 N H_2SO_4 for the acid bottle, 10% KOH for the base bottle, and 50% antifoam B for the antifoam bottle. For the sample port, be sure that the tubing that will come into contact with the culture broth has a small opening to allow the XAD to pass through into the vial for collecting daily samples. Stir the mixture completely before autoclaving to evenly distribute the components. Calibrate the pH probe and test dissolved oxygen probe to ensure proper functioning. Use a small antifoam probe, ~3 inches in length. For the bottles, use tubing that can be sterile welded, but use silicone tubing for the sample port. Make sure all fittings are secure and the tubings are clamped off, not too tightly, with C-clamps. Do not clamp the tubing to the exhaust condenser. Attach 0.2 μm filter disks to any open tubing that is in contact with the air. Use larger ACRO 50 filter disks for larger tubing, such as the exhaust condenser and the air inlet tubing. Prepare a sterile empty bottle for the inoculum. Autoclave at 121°C with a sterilization time of 90 minutes. Once the reactor has been taken out of the

5

autoclave, connect the tubing to the acid, base, and antifoam bottles through their respective pump heads. Release the clamps to these bottles, making sure the tubing has not been welded shut. Attach the temperature probe to the control unit. Allow the reactor to cool, while sparging with air through the air inlet at a low air flow rate.

10

5 After ensuring the pumps are working and there is no problem with flow rate or clogging, connect the hoses from the water bath to the water jacket and to the exhaust condenser. Make sure the water jacket is nearly full. Set the temperature to 32°C. Connect pH, D.O., and antifoam probes to the main control unit. Test the antifoam probe for proper functioning. Adjust the set point of the culture to 7.4. Set the agitation to 400 RPM.

15

10 Calibrate the D.O. probe using air and nitrogen gas. Adjust the airflow using the rate at which the fermentation will operate, e.g. 1 LPM (liter per minute). To control the dissolved oxygen level, adjust the parameters under the cascade setting so that agitation will compensate for lower levels of air to maintain a D.O. value of 50%. Set the minimum and maximum agitation to 400 and 1000 RPM respectively, based on the settings of the control unit. Adjust the settings, if necessary.

20

25

15 Check the seed culture for any contamination before inoculating the fermenter. The *Sorangium cellulosum* cells are rod shaped like a pill, with 2 large distinct circular vacuoles at opposite ends of the cell. Length is approximately 5 times that of the width of the cell. Use a 10% inoculum (minimum) volume, e.g. 400 mL into 4 L of production medium. Take an initial sample from the vessel and check against the bench pH. If the difference between the fermenter pH and the bench pH is off by ≥ 0.1 units, do a 1 point recalibration. Adjust the deadband to 0.1. Take daily 25 mL samples noting fermenter pH, bench pH, temperature, D.O., airflow, agitation, acid, base, and antifoam levels. Adjust pH if necessary. Allow the fermenter to run for seven days before harvesting.

30

20

35

25

40 Extraction and analysis of compounds is performed substantially as described above in Example 4. In brief, fermentation culture is extracted twice with ethyl acetate, and the ethyl acetate extract is concentrated to dryness and dissolved/suspended in ~500 μ L of MeCN-H₂O (1:1). The sample is loaded onto a 0.5 mL Bakerbond ODS SPE cartridge pre-equilibrated with MeCN-H₂O (1:1). The cartridge is washed with 1 mL of the same solvent, followed by 2 mL of MeCN. The MeCN eluent is concentrated to dryness, and the residue is dissolved in 200 μ L of MeCN. Samples (50 μ L) are analyzed by HPLC/MS on a system comprised of a Beckman System Gold HPLC and PE Sciex API100LC single quadrapole MS-based detector equipped with an atmospheric pressure

45

30

50

55

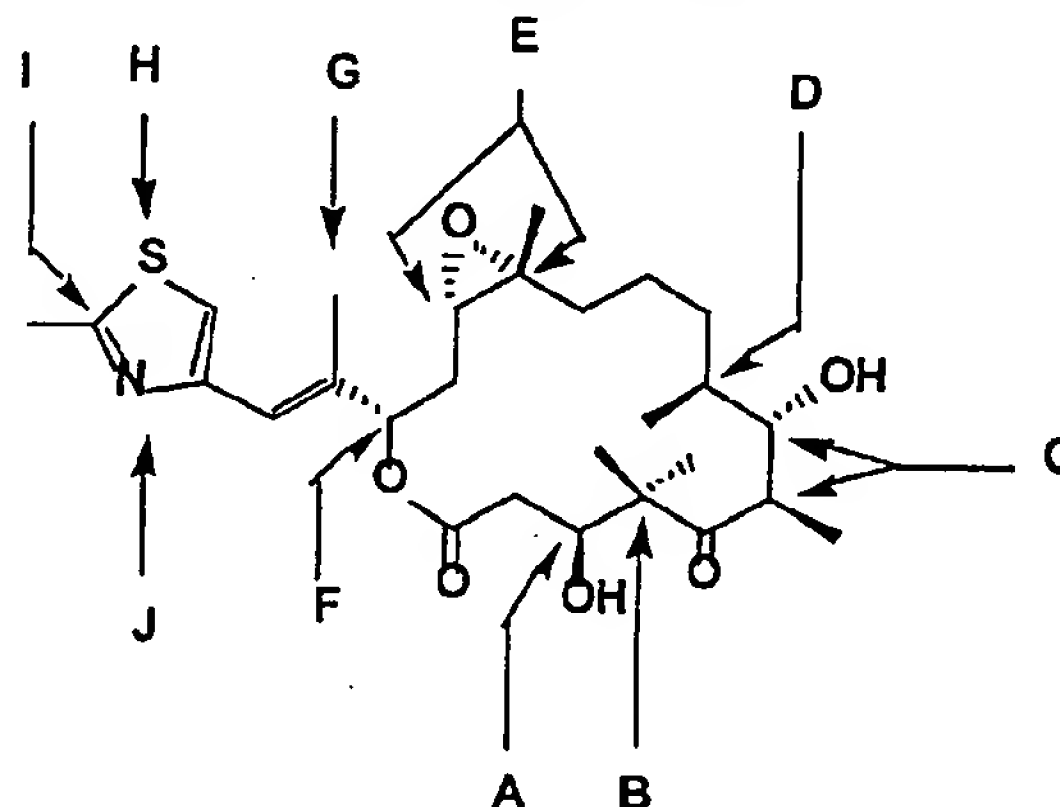
chemical ionization source. Ring and orifice voltages are set to 75V and 300V, respectively, and a dual range mass scan from m/z 290-330 and 450-550 is used. HPLC conditions: Metachem 5 μ ODS-3 Inertsil (4.6 X 150 mm); 100% H₂O for 1 min, then to 100% MeCN over 10 min at 1 mL/min. Epothilone A elutes at 0.2 min under these conditions and gives characteristic ions at m/z 494 (M+H), 476 (M+H-H₂O), 318, and 306.

Example 8

Epothilone Derivatives as Anti-Cancer Agents

The novel epothilone derivatives shown below by Formula (1) set forth above are potent anti-cancer agents and can be used for the treatment of patients with various forms of cancer, including but not limited to breast, ovarian, and lung cancers.

The epothilone structure-activity relationships based on tubulin binding assay are (see Nicolaou *et al.*, 1997, Angew. Chem. Int. Ed. Engl. 36: 2097-2103, incorporated herein by reference) are illustrated by the diagram below.



A) (3S) configuration important; B) 4,4-ethano group not tolerated; C) (6R, 7S) configuration crucial; D) (8S) configuration important, 8,8-dimethyl group not tolerated; E) epoxide not essential for tubulin polymerization activity, but may be important for cytotoxicity; epoxide configuration may be important; R group important; both olefin geometries tolerated; F) (15S) configuration important; G) bulkier group reduces activity; H) oxygen substitution tolerated; I) substitution important; J) heterocycle important.

Thus, this SAR indicates that modification of the C1-C8 segment of the molecule can have strong effects on activity, whereas the remainder of the molecule is relatively

5 tolerant to change. Variation of substituent stereochemistry with the C1-C8 segment, or
removal of the functionality, can lead to significant loss of activity. Epothilone derivative
compounds A-H differ from epothilone by modifications in the less sensitive portion of
10 the molecule and so possess good biological activity and offer better pharmacokinetic
5 characteristics, having improved lipophilic and steric profiles.

These novel derivatives can be prepared by altering the genes involved in the
biosynthesis of epothilone optionally followed by chemical modification. The 9-hydroxy-
15 epothilone derivatives prepared by genetic engineering can be used to generate the
carbonate derivatives (compound D) by treatment with triphosgene or 1,1'
10 carbonyldiimidazole in the presence of a base. In a similar manner, the 9,11-dihydroxy-
epothilone derivative, upon proper protection of the C-7 hydroxyl group if it is present,
20 yields the carbonate derivatives (compound F). Selective oximation of the 9 oxo-
epothilone derivatives with hydroxylamine followed by reduction (Raney nickel in the
presence of hydrogen or sodium cyanoborohydride) yield the 9-amino analogs. Reacting
25 these 9-amino derivatives with p-nitrophenyl chloroformate in the presence of base and
15 subsequently reacting with sodium hydride will produce the carbamate derivatives
(compound E). Similarly, the carbamate compound G, upon proper protection of the C7
30 hydroxyl group if it is present, can be prepared from the 9-amino-11 hydroxy-epothilone
derivatives.

20 Illustrative syntheses are provided below.

Part A. Epothilone D-7, 9-cyclic carbonate

35 To a round bottom flask, a solution of 254 mg epothilone D in 5 mL of methylene
chloride is added. It is cooled by an ice bath, and 0.3 mL of triethyl amine is then added.
To this solution, 104 mg of triphosgene is added. The ice bath is removed, and the mixture
25 is stirred under nitrogen for 5 hours. The solution is diluted with 20 mL of methylene
40 chloride and washed with dilute sodium bicarbonate solution. The organic solution is dried
over magnesium sulfate and filtered. Upon evaporation to dryness, the epothilone D-7, 9 -
cyclic carbonate is isolated.

30 Part B. Epothilone D-7,9-cyclic carbamate

(i) 9-amino-epothilone D

50 To a rounded bottom flask, a solution of 252 mg 9-oxo-epothilone D in 5 mL of
methanol is added. Upon the addition of 0.5 mL 50% hydroxylamine in water and 0.1 mL

5 acetic acid, the mixture is stirred at room temperature overnight. The solvent is then removed under reduced pressure to yield the 9-oxime-epothilone D. To a solution of this 9 oxime compound in 5 mL of tetrahydrofuran (THF) at ice bath is added 0.25 mL 1M
10 solution of cyanoborohydride in THF. After the mixture is allowed to react for 1 hour, the ice bath is removed, and the solution is allowed to warm slowly to room temperature. One mL of acetic acid is added, and the solvent is then removed under reduced pressure. The residue is dissolved in 30 mL of methylene chloride and washed with saturated sodium
15 chloride solution. The organic layer is separated and dried over magnesium sulfate and filtered. Upon evaporation of the solvent yields the 9-amino-epothilone D.

10 (ii) Epothilone D-7,9-cyclic carbamate

20 To a solution of 250 mg of 9-amino-epothilone D in 5 mL of methylene is added 110 mg of 4-nitrophenyl chloroformate followed by the addition of 1 mL of triethylamine. The solution is stirred at room temperature for 16 hours. It is diluted with 25 mL of methylene chloride. The solution is washed with saturated sodium chloride and the organic
25 layer is separated and dried over magnesium sulfate. After filtration, the solution is evaporated to dryness at reduced pressure. The residue is dissolved in 10 mL of dry THF. Sodium hydride, 40 mg (60% dispersion in mineral oil), is added to the solution in an ice bath. The ice bath is removed, and the mixture is stirred for 16 hours. One-half mL of
30 acetic acid is added, and the solution is evaporated to dryness under reduced pressure. The residue is re-dissolved in 50 mL methylene chloride and washed with saturated sodium chloride solution. The organic layer is dried over magnesium sulfate and the solution is
35 filtered and the organic solvent is evaporated to dryness under reduced pressure. Upon purification on silica gel column, the epothilone D-7,9-carbamate is isolated.

25 The invention having now been described by way of written description and
40 examples, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments and that the foregoing description and examples are for purposes of illustration and not limitation of the following claims.

Claims

5

10

15

20

25

30

35

40

45

50

55

5

Claims

10

1. An isolated recombinant nucleic acid compound that comprises a nucleotide sequence encoding at least a domain of an epothilone polyketide synthase (PKS) protein and/or encoding a functional region of an epothilone modification enzyme.

15

2. The nucleic acid of claim 1, wherein said domain is selected from the group consisting of a loading domain, a thioesterase domain, an NRPS, an AT domain, a KS domain, an ACP domain, a KR domain, a DH domain, and an ER domain, a methyl transferase domain and a functional oxidase domain.

20

3. The nucleic acid of claim 1 or 2 that comprises the coding sequence of an *epoA* gene, and/or

25

15

- the coding sequence of an *epoB* gene, and/or
the coding sequence of an *epoC* gene, and/or
the coding sequence of an *epoD* gene, and/or
the coding sequence of an *epoE* gene, and/or
the coding sequence of an *epoF* gene, and/or
the coding sequence of an *epoK* gene, and/or
the coding sequence of an *epoL* gene.

30

20

35

4. The nucleic acid of any of claims 1-3 that further comprises a promoter positioned to transcribe said encoding nucleotide sequence in host cells in which said promoter is operable.

25

40

5. The nucleic acid of claim 4, wherein said promoter is a promoter from a *Sorangium* gene, or

45

30

- from a *Myxococcus* gene, or
from a *Streptomyces* gene, or
from an epothilone PKS gene, or
from a *pilA* gene, or
from an actinorhodin PKS gene.

50

55

5

6. The nucleic acid of any of claims 1-5 that is a recombinant DNA expression vector.

10

5

7. Host cells which contain the nucleic acid of any of claims 4-6.

15

8. The cells of claim 7 which are *Sorangium* cells, or
Myxococcus cells, or
Pseudomonas cells, or
Streptomyces cells.

10

20

9. A method to produce a polyketide which method comprises culturing the cells of claim 7 or 8 under conditions wherein the encoding nucleotide sequence is expressed to obtain a functional PKS.

25

15

10. A recombinant *Sorangium cellulosum* host cell that contains a mutated gene for an epothilone PKS protein or epothilone modification enzyme, wherein said mutated gene was inserted in whole or in part into genomic DNA of said cell by homologous recombination with a recombinant vector comprising all or a part of an epothilone PKS gene or epothilone modification gene.

30

20

35

11. The recombinant host cell of claim 10 that
makes epothilone C or D but not A or B due to a mutation inactivating or deleting an *epoK* gene, or
makes epothilone A or C but not B or D due to a mutation in *epoD* altering module
4 AT domain specificity, or
makes epothilone B or D but not A or C due to a mutation in *epoD* altering module
4 AT domain specificity, or
makes epothilone C but not epothilone A, B or D due to a mutation in *epoD*
altering module 4 AT domain specificity and a mutation in *epoK*, or
makes epothilone D but not epothilone A, B or C due to a mutation in *epoD*
altering module 4 AT domain specificity and a mutation in *epoK*.

25

40

45

30

50

55

5

10

5

12. Recombinant *Streptomyces* or *Myxococcus* host cells that express an epothilone PKS gene or an epothilone modification enzyme gene, optionally comprising one or more of said epothilone PKS or modification enzyme genes integrated into their chromosomal DNA and/or one or more of said epothilone PKS or modification enzyme genes on an extrachromosomal expression vector.

15

13. The host cells of claim 12 or 13 that are *S. coelicolor* CH999.

14. A method to produce an epothilone or epothilone derivative which comprises culturing the cells of claims 12 or 13.

20

15. A modified functional epothilone PKS wherein said modification comprises at least one of:

25

15

replacement of at least one AT domain with an AT domain of different specificity;
inactivation of the NRPS-like module 1 or of the KS2 catalytic domain;
inactivation of at least one activity in at least one β -carbonyl modification domain;
addition of at least one of KR, DII and ER activity in at least one β -carbonyl modification domain; and
replacement of the NRPS module 1 with an NRPS of different specificity.

30

20

35

16. The modified PKS of claim 15 contained in a cell or contained in a cell-free system, wherein said cell or system contains additional enzymes for modification of the product of said epothilone PKS.

40

25

17. The modified PKS of claim 16 wherein said modifying enzymes comprise at least one of a methyltransferase, an oxidase or a glycosylation enzyme.

45

30

18. A method to prepare an epothilone derivative which method comprises providing substrates including extender units to the modified PKS of any of claims 15-17.

50

19. A modified functional epothilone PKS wherein said modification comprises inactivation of the NRPS of module 1 or the KS2 of module 2 thereof.

55

5

20. A method to make an epothilone derivative which method comprises contacting the modified PKS of claim 19 with a module 2 substrate or a module 3 substrate and extender units.

10

5 21. Recombinant host cells which comprise the modified PKS of any of claims 15-17 or 19.

15

22. The cells of claim 21 that produce an epothilone derivative selected from the group consisting of 16-desmethyl epothilones, 14-methyl epothilones, 11-hydroxyl epothilones, 10-methyl epothilones, 8,9-anhydro epothilones, 9-hydroxyl epothilones, 9-keto epothilones, 8-desmethyl epothilones, and 6-desmethyl epothilones.

20

23. A compound selected from the group consisting of 16-desmethyl epothilones, 14-methyl epothilones, 11-hydroxyl epothilones, 10-methyl epothilones, 8,9-anhydro epothilones, 9-hydroxyl epothilones, 9-keto epothilones, 8-desmethyl epothilones, and 6-desmethyl epothilones.

25

24. A recombinant PKS enzyme that comprises one or more domains, modules, or proteins of a non-epothilone PKS and one or more domains, modules, or proteins of an epothilone PKS, and/or contains a loading domain that comprises a KS^Q domain.

30

35

25. The PKS enzyme of claim 24, wherein said PKS comprises a DEBS loading domain and 5 modules of DEBS and an NRPS of the epothilone PKS,

25

40

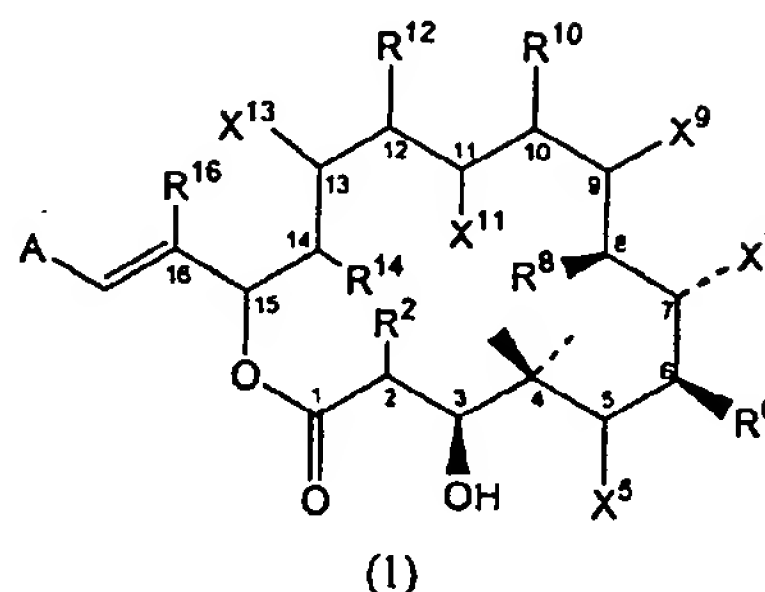
wherein said PKS comprises all of a non-epothilone PKS with an MT domain of the epothilone PKS

45

50

55

26. A compound of the formula:



including the glycosylated forms thereof and stereoisomeric forms where the stereochemistry is not shown,

wherein A is a substituted or unsubstituted straight, branched chain or cyclic alkyl, alkenyl or alkynyl residue optionally containing 1-3 heteroatoms selected from O, S and N; or wherein A comprises a substituted or unsubstituted aromatic residue;

R² represents H, H, or H, lower alkyl, or lower alkyl, lower alkyl;

X⁵ represents =O or a derivative thereof, or H, OH or H, NR₂ wherein R is H, alkyl or acyl, or H, OCOR₂, H, OCONR₂ wherein R is H or alkyl, or is H, H;

R⁶ represents H or lower alkyl, and the remaining substituent on the corresponding carbon is H;

X⁷ represents OR, or NR₂, wherein R is H, alkyl or acyl or is OCOR, or OCONR₂ wherein R is H or alkyl or X⁷ taken together with X⁹ forms a carbonate or carbamate cycle, and wherein the remaining substituent on the corresponding carbon is H;

R⁸ represents H or lower alkyl and the remaining substituent on the carbon is H;

X⁹ represents =O or a derivative thereof, or H, OR or H, NR₂ wherein R is H, alkyl or acyl, or is H, OCOR or H, OCONR₂, wherein R is H or alkyl, or represents H, H or wherein X⁹ together with X⁷ or with X¹¹ can form a cyclic carbonate or carbamate;

R¹⁰ is H, H, or H, lower alkyl, or lower alkyl, lower alkyl;

X¹¹ is =O or a derivative thereof, or H, OR, or H, NR₂ wherein R is H, alkyl or acyl or H, OCOR or H, OCONR₂ wherein R is H or alkyl, or is H, H or wherein X¹¹ in combination with X⁹ may form a cyclic carbonate or carbamate;

R¹² is H, H, or H, lower alkyl, or lower alkyl, lower alkyl;

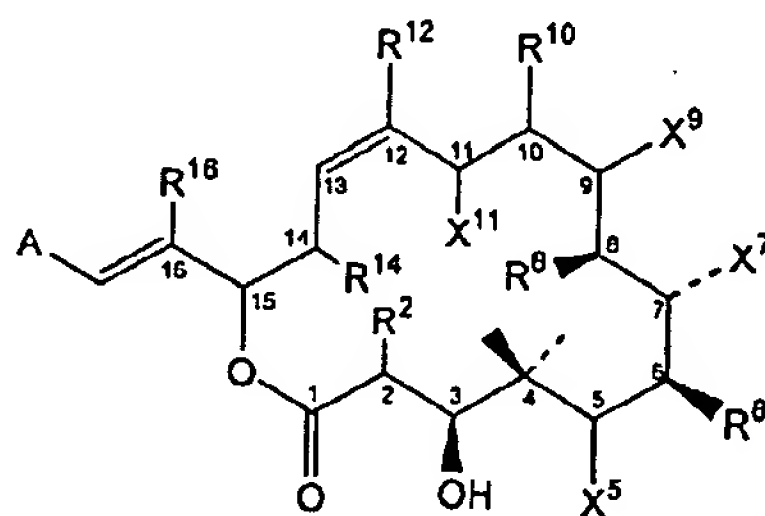
X¹³ is =O or a derivative thereof, or H, OR or H, NR₂ wherein R is H, alkyl or acyl or is H, OCOR or H, OCONR₂ wherein R is H or alkyl;

R^{14} is H, H, or H, lower alkyl, or lower alkyl, lower alkyl;

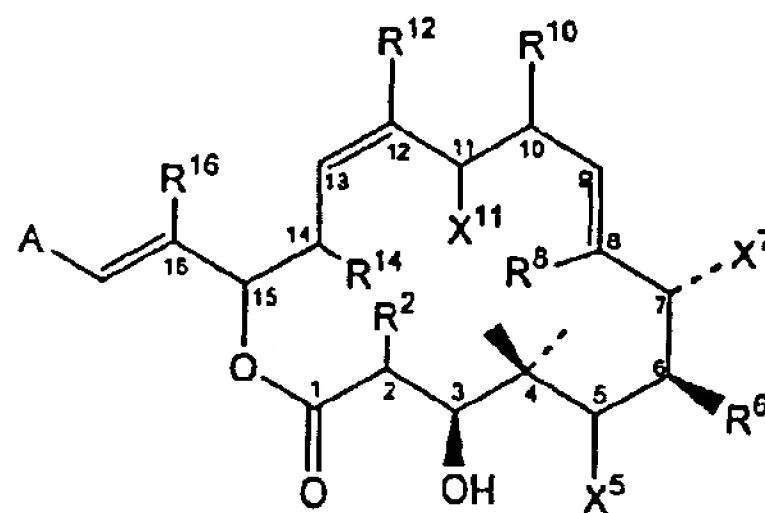
R^{16} is H or lower alkyl; and

wherein optionally H or another substituent may be removed from positions 12 and 13 and/or 8 and 9 to form a double bond, wherein said double bond may optionally be converted to an epoxide.

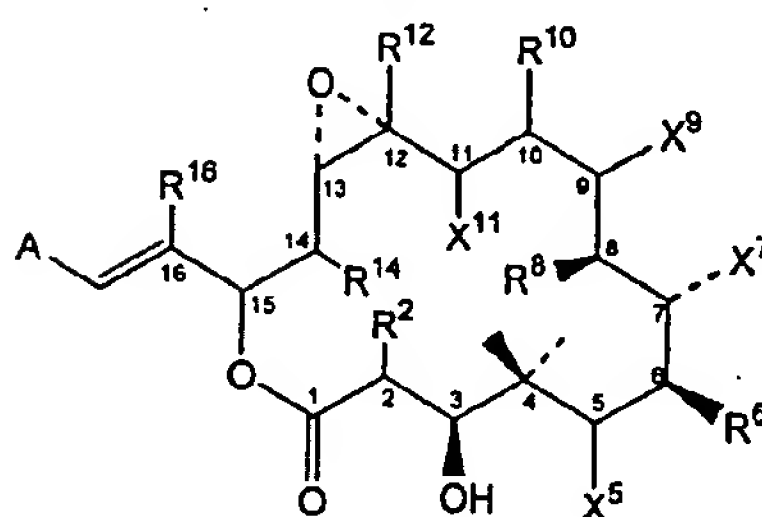
27. A compound of the formula



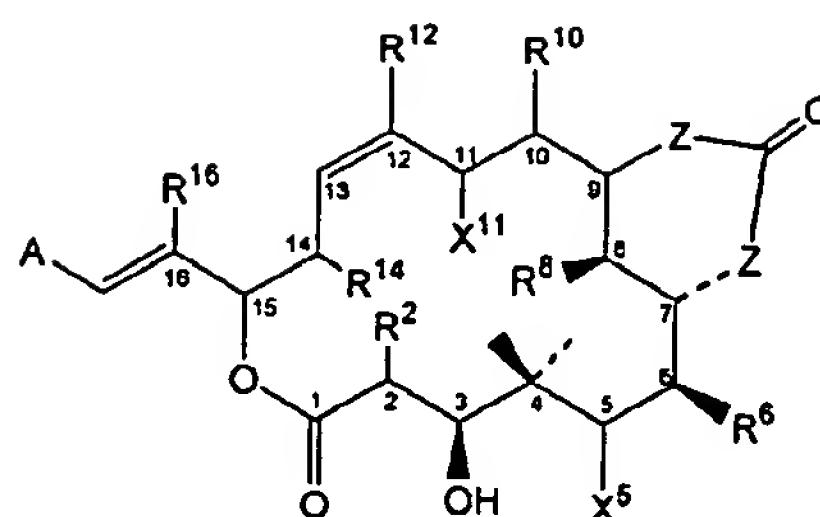
1(a),



1(b),

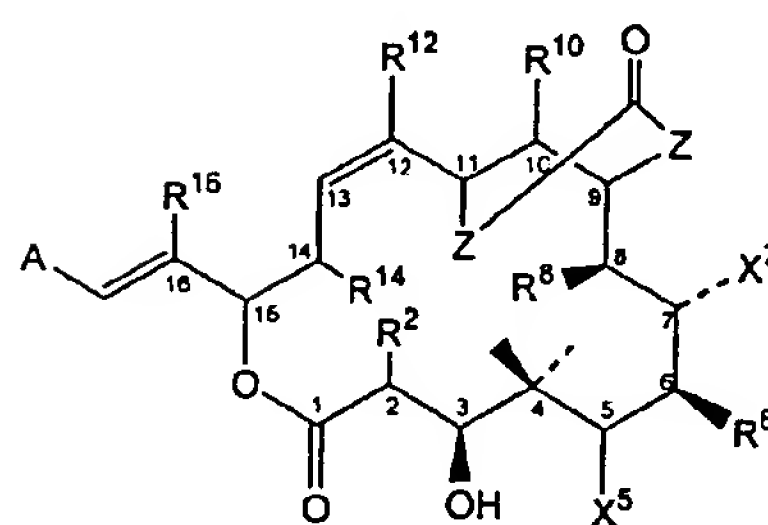


1(c),



1(d),

and



1(e)

wherein both Z are O or one Z is N and the other Z is O and the remaining substituents are defined as in claim 26.

28. A recombinant vector selected from the group consisting of pKOS35-70.8A3, pKOS35-70.1A2, pKOS35-70.4, pKOS35-79.85, pKOS039-124R, and pKOS039-126R.